

Національна академія наук України
Міністерство освіти і науки України
Міжнародний науково-навчальний центр
інформаційних технологій і систем

НЕСОЛОВА МАРІНА МИКОЛАЇВНА

УДК 004[616.37–001 + 616.366 + 616.831]

**МОДЕЛІ І МЕТОДИ ОЦІНКИ СТУПЕНЯ ТЯЖКОСТІ СТАНУ
ПАЦІЄНТІВ ДЛЯ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ
ЛІКАРЯ**

05.13.09 – медична і біологічна інформатика та кібернетика

Автореферат дисертації на здобуття наукового ступеня
кандидата технічних наук

Дисертацією є рукопис

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки України

Науковий керівник: доктор біологічних наук, професор
Кочина Марина Леонідівна
Харківська медична академія післядипломної освіти,
професор кафедри клінічної інформатики та інформаційних технологій в управлінні охороною здоров'я.

Офіційні опоненти:

доктор технічних наук, професор,
Злепко Сергій Макарович,
Вінницький національний технічний університет,
зав. кафедри проектування медико-біологічної апаратури;

кандидат технічних наук, доцент,
Литвинов Олександр Анатолійович,
Дніпропетровський національний університет,
доцент кафедри електронних обчислювальних машин.

Захист відбудеться « 29 » квітня 2015 р. о 12-00 годині на засіданні спеціалізованої вченої ради Д 26.171.03 в Міжнародному науково-навчальному центрі інформаційних технологій і систем НАН і МОН України за адресою: м. Київ, 03680 МСП, пр. Академіка Глушкова, 40, корпус № 11.

З дисертацією можна ознайомитись у бібліотеці Міжнародного науково-навчального центру інформаційних технологій і систем НАН і МОН України за адресою: пр. Академіка Глушкова, 40, корпус № 11, м. Київ, 03680 МСП, Україна.

Автореферат розісланий «27» березня 2015 р.

Вчений секретар
спеціалізованої вченої ради Д 26.171.03

Т.М. Гонтар

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Первинний огляд пацієнта при надходженні у стаціонар починається з оцінювання ступеня тяжкості стану (СТС). Таке оцінювання доводиться виконувати в подальшому неодноразово для обґрунтування вибору лікувальної тактики, визначення ефективності лікувальних заходів і т. ін. Правильна оцінка СТС пацієнтів дозволяє коректно прогнозувати подальший перебіг і клінічний наслідок захворювання, формувати однорідні групи пацієнтів для наукових досліджень, ефективно планувати ресурси охорони здоров'я, а особлива увага приділяється їй у страховій медицині (Потаніна О.К., 2012; Євтюков Г.М., 2006; Купцов А.А., 2001; Вильдеман О.В., 2010). Сучасним способом оцінювання СТС пацієнтів і постраждалих із травмами є застосування інтегральних шкал і систем, що ґрунтуються на математичних моделях, які розробляються на базі аналізу накопичених даних про стан пацієнтів внаслідок різних захворювань, травматичних чи інших ушкоджень органів і систем організму (Гаїн Ю.М., 2005; Потаніна О.К., 2012; Євтюков Г.М., 2006; Садчиков Д.В., 2011; Березка М.І., 2013). Сучасний підхід до прогнозування подальшого перебігу та клінічного наслідку захворювань, а також до оцінювання СТС пацієнтів базується на аналізі причинно-наслідкових зв'язків між характеристиками стану, що спостерігаються, та формуванні на основі результатів цього аналізу вирішальних правил, відповідно до яких здійснюється процес діагностики. Оскільки величини, що містяться в масивах медичних даних у більшості своїй є стохастичними, а досліджувані явища та процеси, як правило, протікають в умовах багатофакторності, одним з найбільш перспективних підходів до побудови математичних моделей у медичних застосуваннях вважається використання методів багатовимірного статистичного аналізу (Халафян О.А., 2010; Ахутін В.М., 2002; Поворознюк О.І., 2009; Вильдеман О.В., 2010). Формалізація оцінювання СТС пацієнта можлива через дослідження на основі аналізу існуючих прецедентів закономірностей, які пов'язують характеристики пацієнта з групою-ступенем тяжкості, до якої він належить. Тому при розробленні методів оцінки СТС пацієнтів одним з можливих підходів є розв'язок вказаної задачі у постановці задачі класифікації з навчанням (навчання із вчителем, або задачі дискримінації). Розробленням математичних моделей, інформаційних технологій і систем для розв'язання задач класифікації зі вчителем у сучасних медичних застосуваннях займалися такі вчені як Гельфанд І.М., Шифрін М.А., Орлов О.І., Халафян О.А., Леонов В.П., Воронцов К.В.

Тим не менш, протягом останніх 30 років справедливою вважається думка (Champion H.R., 1980; Kaukinen L., 1984) про те, що математичні моделі для оцінювання СТС пацієнтів, які формалізують та забезпечують підтримку прийняття рішень лікаря, повинні розроблятися і корегуватися не тільки з урахуванням специфіки кожного конкретного захворювання чи стану, а також із плином часу, в залежності від регіону та інших факторів. Крім того, специфіка вхідних даних у галузях людської діяльності, що складно формалізуються, до яких належить і медицина, накладає додаткові вимоги та вносить свої особливості до підходів до побудови моделей для вирішення завдань даної галузі, які в першу чергу повинні враховувати можливу неповність інформації та дозволяти сумісний аналіз множини різнотипних ознак.

Важливість правильного та своєчасного оцінювання стану пацієнта, прогнозування подальшого перебігу та клінічного наслідку захворювань обумовлює актуальність та необхідність

розроблення нових моделей і методів оцінювання ступеня тяжкості стану пацієнтів для забезпечення підтримки прийняття рішень лікаря.

Зв'язок роботи з науковими програмами й темами. Дослідження, результати яких викладені в дисертації, проводились у відповідності до договорів (№ 560 / 12-12 від 12.12.2012 р., 2012—2018 рр.) про науково-практичне співробітництво, які виконувалися у Національному фармацевтичному університеті, Харківському національному медичному університеті та обласній клінічній лікарні м. Харкова та були укладені в рамках виконання науково-дослідних робіт ДУ «Інститут загальної і невідкладної хірургії АМН України» – ВН.3.05. «Розробити нові способи гемостазу, лікування та профілактики ускладнень синдрому внутрішньочеревної гіпертензії при травмі органів живота та кровотечах» (державний реєстраційний № 0105U000897) і ВН.1.08 «Розробити діагностичну тактику і нові способи хірургічного лікування травматичних ушкоджень органів травного тракту» (державний реєстраційний № 0108U001029).

Мета дослідження: підвищення точності оцінки ступеня тяжкості стану пацієнтів за рахунок розроблення і використання методів і моделей класифікації з навчанням та їх реалізації в інформаційній технології оцінювання ступеня тяжкості стану пацієнтів.

Задачі дослідження:

1. Провести порівняльний аналіз методів і моделей, які можуть бути застосовані для формалізації процесу оцінювання ступеня тяжкості стану пацієнтів.
2. Розробити методи класифікації з навчанням та створення композицій класифікаторів для оцінювання ступеня тяжкості стану пацієнтів.
3. Розробити математичні моделі прогнозування клінічного наслідку і оцінювання ступеня тяжкості стану пацієнтів, а також програмне забезпечення, що реалізує побудовані моделі.
4. Розробити інформаційну технологію оцінювання ступеня тяжкості стану пацієнтів і провести експериментальне дослідження даної технології на клінічних даних пацієнтів із травматичними ушкодженнями підшлункової залози, травматичним панкреатитом, захворюваннями жовчовивідних проток та інсультами.

Об'єкт дослідження: процес оцінювання ступеня тяжкості стану пацієнта.

Предмет дослідження: інформаційна технологія, математичні моделі і методи оцінювання ступеня тяжкості стану пацієнтів.

Методи дослідження. При розробленні методу побудови класифікаторів за навчальною інформацією використовувався статистичний підхід до розв'язання задачі класифікації, а саме: його побудова базувалася на аналізі результатів застосування методів геометричної інтерпретації багатовимірних даних (кореспондентського аналізу і багатовимірного шкалування) з наступною побудовою класифікатора у вигляді алгоритму обчислення оцінок. При розробленні способів формування композицій класифікаторів використовувались принципи спеціалізації та зваженого голосування. Побудова математичних моделей базувалась на методах математично-статистичного моделювання. При розробленні програмного забезпечення застосовувались методи об'єктно-орієнтованого програмування.

Наукова новизна отриманих результатів.

Отримали подальший розвиток методи розв'язання задач класифікації з учителем за рахунок використання аналізу структури підмножин ознак у просторі об'єктів з урахуванням додаткової ознаки у вигляді вектора міток класів, на відміну від традиційного представлення класів

як підмножин об'єктів у просторі ознак, що дозволяє застосовувати методи геометричної інтерпретації структури багатовимірних даних для побудови класифікаторів як алгоритмів обчислення оцінок стану досліджуваних об'єктів.

Вперше запропоновано та формалізовано метод кількісної оцінки сили впливу ознак на приналежність об'єктів до класів, який базується на аналізі просторової структури взаємозв'язків між ознаками-предикторами й класами, що надає можливість визначати коефіцієнти оціночних функцій класифікації для врахування нелінійності та немонотонності зміни оцінки приналежності об'єкта до класу в залежності від предикторних змінних.

Отримали подальший розвиток методи побудови композицій класифікаторів на основі комбінування принципів спеціалізації та зваженого голосування, що дозволяє підвищити якість класифікації у разі сумісного використання декількох моделей.

Вперше побудовано математичні моделі оцінювання ступеня тяжкості стану і прогнозування клінічного наслідку, що базуються на використанні розроблених методів побудови класифікаторів та їхніх композицій та дозволяють класифікувати пацієнтів із *травмою підшлункової залози* на три ступеня тяжкості стану (із загальною точністю 87,3%, специфічністю 88,2% і чутливістю 89,7%) та в залежності від клінічного наслідку (із загальною точністю 86,9%, специфічністю 85,9% і чутливістю 87,8%); проводити диференційну діагностику *захворювань жовчовивідних проток* (із загальною точністю 80%, специфічністю 100% і чутливістю 75%); визначати клінічний наслідок *інсульту* (із загальною точністю 91,9%, специфічністю 96% і чутливістю 76%).

Вперше розроблено інформаційну технологію, яка реалізує побудовані математичні моделі оцінювання ступеня тяжкості стану пацієнтів із травматичними пошкодженнями підшлункової залози в автоматизованій системі, яка дозволяє класифікувати постраждалих при надходженні в стаціонар і в процесі лікування. Проведено експериментальне дослідження даної інформаційної технології на клінічних даних пацієнтів із травмами підшлункової залози, травматичним панкреатитом, інсультом, захворюваннями жовчовивідних проток, результати якого вказали на підвищення точності оцінки стану пацієнтів (за показниками загальної точності класифікації на $4,4 \div 8,9\%$, специфічності – на $4,2 \div 6\%$, чутливості – на $7,6 \div 24,8\%$) у порівнянні з класичними методами статистичного аналізу.

Практичне значення отриманих результатів.

Розроблена інформаційна технологія, яка реалізує математичні моделі, побудовані на основі розроблених методів класифікації з навчанням та формування композицій класифікаторів, дозволяє підвищити точність оцінювання СТС та прогнозування клінічного наслідку пацієнтів, які є підставою для вибору тактики лікування та передбачення подальшого перебігу захворювання, що дає змогу покращити ефективність лікувальних заходів.

Розроблений метод побудови класифікаторів на основі геометричної інтерпретації структури багатовимірних даних дозволяє отримувати моделі прогнозування клінічного наслідку й оцінювання СТС пацієнтів, а також розв'язки інших прикладних задач класифікації з навчанням.

Вдосконалені способи побудови композицій класифікаторів, які поєднують принципи спеціалізації та зваженого голосування, можуть застосовуватись для формування ансамблів моделей побудованих різними методами, що дозволяє підвищити точність класифікації у практичних застосуваннях.

Метод оцінювання стану підшлункової залози при механічних ушкодженнях, що розроблено в

рамках досліджень (патент “Спосіб оцінки стану підшлункової залози при механічних пошкодженнях” [9]), може застосовуватися при оцінюванні СТС пацієнтів із травмою підшлункової залози в залежності від стану ушкодженого органу.

Результати дисертаційної роботи впроваджені в наукові дослідження та навчальний процес кафедри клінічної інформатики та інформаційних технологій в управлінні охороною здоров'я Харківської медичної академії післядипломної освіти, кафедри фармакоінформатики Національного фармацевтичного університету, кафедри радіоелектронних пристроїв та кафедри основ радіотехніки Харківського національного університету радіоелектроніки, кафедри хірургії № 2 Харківського національного медичного університету, в лікувальну роботу клініки хірургії Обласної клінічної лікарні м. Харкова.

Особистий внесок здобувача. Автор дисертаційної роботи розробив інформаційну технологію оцінювання СТС пацієнтів із травматичними пошкодженнями підшлункової залози, розробив метод розв'язання задачі класифікації з навчанням, заснований на геометричній інтерпретації структури багатовимірних даних, розробив моделі для оцінювання СТС та прогнозування клінічного наслідку у випадку травм підшлункової залози, захворювань жовчовивідних проток та інсультів і провів їхнє експериментальне дослідження на клінічних даних.

У роботах зі співавторами авторові належить: [1] – ідея та реалізація методу оцінювання СТС пацієнтів; [2] – проведення оцінювання впливу ступеня травматичного ушкодження на СТС пацієнтів; [3, 4, 6] – ідея та реалізація методів експрес-оцінки СТС при травмі підшлункової залози, заснованих на застосуванні багатовимірного шкалування та комбінації дискримінантних функцій; [5] – статистичне обґрунтування спрощеного бального методу оцінювання СТС при травмі підшлункової залози; [7] – розроблення й апробація на клінічних даних математичних моделей оцінювання СТС при травмі підшлункової залози, що є складовими інформаційної технології; [8] – ідея та реалізація розроблення методу побудови класифікаторів за навчальною інформацією, а також вдосконалення способів створення їхніх композицій.

Апробація результатів роботи. Матеріали дисертаційної роботи доповідалися та обговорювалися на всеукраїнських і міжнародних науково-практичних телеконференціях «Актуальні питання дистанційної освіти та телемедицини» (Запоріжжя—Київ, 2009, 2012, 2013, 2014), Всеукраїнському конгресі «Сьогодення та майбутнє фармації» (Харків, 2008), 3-му Міжнародному форумі «Прикладная радиоэлектроника. Состояние и перспективы развития» (Харків, 2008), Всеукраїнських науково-практичних конференціях студентів і молодих вчених «Актуальні питання створення нових ЛЗ» (Харків, 2010, 2011, 2013), конференції з міжнародною участю «Медична та біологічна інформатика і кібернетика: віхи розвитку» (Київ, 2011), міжнародній конференції «Информационные технологии в кардиологии» (Харків, 2013), науково-практичній конференції з міжнародною участю «Информационные технологии в неврологии, психиатрии, эпилептологии и медицинской статистике» (Київ, 2013).

Публікації. Основні результати дисертаційних досліджень опубліковано в 20 друкованих наукових працях, з яких 8 статей у фахових журналах і збірниках, що входять до переліку ДАК України, 2 статті у інших виданнях, 1 патент і 9 тез у збірниках матеріалів та тез доповідей на наукових конференціях.

Структура і обсяг дисертації. Дисертація складається із вступу, чотирьох розділів, висновків,

списку використаних джерел із 215 найменувань (23 сторінки), 7 додатків (11 сторінок). Загальний обсяг дисертаційної роботи складає 188 сторінок, з них 150 сторінок основного друкованого тексту, ілюстровано 40 рисунками і 10 таблицями.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність досліджень, визначено їхню мету, предмет і об'єкт, сформульовано задачі дослідження і методи їх розв'язання; визначено наукову новизну і практичне значення отриманих результатів.

У першому розділі проведено огляд літературних джерел, що стосуються способів оцінювання СТС пацієнтів. Сформульовано постановку задачі оцінювання СТС пацієнта як задачі класифікації з навчанням. Проведено аналіз методів, що використовуються в медичних застосуваннях для розв'язку задач класифікації з навчанням, у результаті якого як найбільш перспективні визначено методи KDD (Knowledge Discovery in Databases) та його основного елементу Data Mining (видобутку, розкопки даних). Серед останніх виявлено цілий пласт методів – візуалізації та графічної (геометричної) інтерпретації багатовимірних даних, – перспективи, потенціал і результати застосування яких у медичних застосуваннях ще недостатньо вивчені.

У другому розділі викладено методичні основи розроблення інформаційної технології оцінювання СТС пацієнтів. Визначено етапи цього процесу, сформульовані задачі етапів та обрані методи їх вирішення (рис. 1).

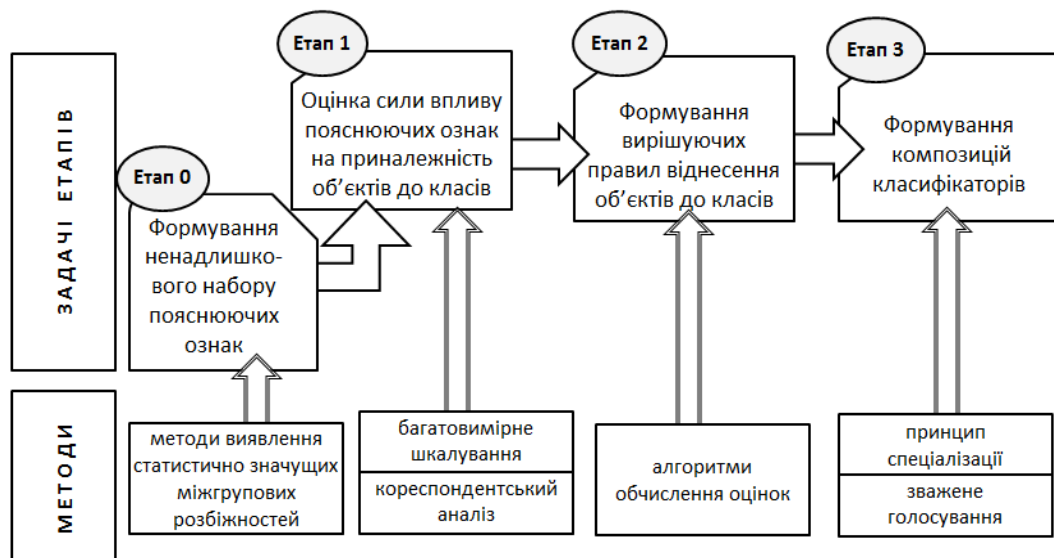


Рис. 1. Етапи розроблення інформаційної технології оцінювання СТС пацієнтів, їх задачі і методи розв'язку.

Аналіз методів графічної інтерпретації структури даних дозволив виділити метод множинного кореспондентського аналізу (Burt, 1950; Greenacre, 1984) як найбільш перспективний для використання при розробленні методу побудови класифікаторів за навчальною інформацією. Він дозволяє використовувати ознаки, виміряні в найпростішій шкалі – шкалі найменувань, та не накладає додаткових вимог на їхні типи даних і закони розподілу, і отримувати при цьому представлення взаємозв'язків ознак із класами об'єктів, що достатньо легко інтерпретуються, у вигляді карт у просторі невеликої вимірності. Для визначення ознак, які мають значущий вплив на

приналежність об'єктів до класів, вирішено використовувати методи множинних групових порівнянь, адекватні типу даних і шкалі вимірювання кожної конкретної ознаки.

Етап формування вирішальних правил класифікації здійснено за допомогою створення моделей алгоритмів типу обчислення оцінок (Журавльов, 1978), в яких: (а) система опорних множин алгоритму визначається на попередньому етапі (етапі 0, рис. 1) розроблення інформаційної технології оцінювання СТС пацієнта; (б) етап 1 визначає способи оцінювання близькості об'єктів і знаходження оцінок за класами, які здійснюються на основі геометричної інтерпретації структури даних; (с) вирішальне правило моделі алгоритму обчислення оцінок, використане в дисертації на етапі 2 розроблення інформаційної технології оцінювання СТС пацієнта, – операція вибору максимальної оцінки.

Наявність етапу 3 (рис. 1) розв'язання основної задачі дисертаційної роботи була обумовлена необхідністю підвищення точності класифікації, яке досягалося за рахунок побудови композицій, які об'єднують декілька моделей класифікаторів. Комбінування класифікаторів вважається найбільш сильним і перспективним прийомом у випадках, коли при розв'язанні практичних задач ні один зі стандартних підходів (вибір моделі з іншого сімейства алгоритмів, або використання альтернативних методів настроювання її параметрів) не дозволяє отримати класифікатор прийнятної точності, або якщо різними методами вдається побудувати декілька вданих класифікаторів зі співмірними характеристиками якості. Проведений аналіз наявних способів формування композицій класифікаторів дозволив визначити евристики, на основі комбінування яких можливе розроблення нових методів побудови композицій. Найбільш перспективними у даній роботі виявилися комбінації методів зваженого голосування, голосування по старшинству і сумішей експертів.

У **третьому розділі** описуються етапи розроблення інформаційної технології оцінювання СТС пацієнтів та її складових: методу побудови класифікаторів на основі геометричної інтерпретації структури багатовимірних даних і методів складання композицій класифікаторів, що названі “рейтинговим голосуванням” і “рейтинговим голосуванням по старшинству”, а також загальні принципи розроблення програмного забезпечення для інформаційної технології.

Метод побудови класифікаторів за навчальною інформацією на основі геометричної інтерпретації структури багатовимірних даних є основною зі складових інформаційної технології, що пропонується. Схематично алгоритм практичного застосування розробленого методу класифікації з навчанням представлено на рис. 2. На даній схемі вхідна навчальна інформація (база даних) відповідає поданню у вигляді матриці ознакового опису об'єктів. На першому етапі відбувається структуризація і перетворення таблиці навчання, які включають і попередню обробку даних. Матриця об'єктів-ознак скорочується шляхом виключення строк, що відповідають спостереженням-викидам, і стовпців, що відповідають колінеарним (залежним) змінним. Також виключаються стовпці матриці, що відповідають показникам, які не змінюються при переході від класу до класу. Такі ознаки вважаються неінформативними для визначення приналежності об'єктів до класів в силу відсутності впливу ознаки «клас» на їхні значення. Кроки із виключення колінеарних і неінформативних предикторів можуть виконуватися в будь-якому порядку. У підсумку виконання зазначених процедур буде отримано ненадлишковий набір n незалежних показників, що є потенціальними предикторами, які обумовлюють приналежність об'єктів до різних класів.

Категоризація значень предикторних змінних відбувається шляхом розбивання областей значень кількісних показників на інтервали, що характерні для різних класів, а для якісних показників – об'єднанням тих категорій їхніх значень, які характерні для одного класу. В результаті для кожного класу буде сформовано набір т.з. еталонів (шаблонів, паттернів), що представляють собою елементарні правила порівняння показників з пороговими значеннями (для кількісних ознак, що описують об'єкти) або перелік категорій, які характерні для даного класу (для якісних ознак). Якщо ознаки представляються як функції на множині об'єктів $x \in X : f_i : X \rightarrow D_{f_i}$ ($i = 1, \dots, n$), то після категоризації $f_i(x)$ буде задаватися як вектор $f_i(x) = (\mu_{g_{il}}(x))_{l=1}^{k_i}$, де $\mu_{g_{il}}(x)$ – характеристичні функції підмножин-категорій (інтервалів) значень ознаки $g_{il} \subset D_{f_i}$, k_i – число підмножин, на які розбивається область значень ознаки D_{f_i} ($1 < k_i \leq m$), m – кількість класів.



Рис. 2. Схема практичного застосування розробленого методу класифікації з навчанням.

Ключовим моментом у розробленні запропонованого методу класифікації з навчанням є робота з класом не як з підмножиною навчальної вибірки, а як з ще однією якісною змінною, що складає ознаковий опис об'єкту. Такий підхід дає можливість розглядати і пояснюючі змінні, і класи, як однорідні показники (виміряні за однаковою шкалою). Таким чином, результатом першого етапу є навчальна матриця «об'єкт–ознака», яка містить $n+1$ стовпець, що відповідає класам об'єктів і категоризованим значенням n незалежних предикторів. Дана матриця являє собою вхідну інформацію для отримання геометричної інтерпретації взаємозв'язків між показниками за допомогою методу множинного кореспондентського аналізу (Burt, 1950; Greenacre, 1984). Якщо навчальну інформацію задано у вигляді таблиці попарних порівнянь, то етап попередньої обробки даних, включаючи приведення змінних до єдиної шкали, може бути опущений. В цьому випадку на етапі формування просторової структури взаємозв'язків ознак найбільш доцільним буде використання методів багатовимірної шкалування (Torgerson, 1952; Kruskal, 1964).

На третьому етапі аналізується отримана просторова структура і обирається метрика, за допомогою якої оцінюються відстані між ознаками на карті, що одержано. Вагові коефіцієнти категорій значень предикторних змінних при прогнозуванні кожного з класів обчислюються як величини обернено пропорційні до відстаней від точок, що їх представляють, до точок, які відповідають класам об'єктів, за формулою:

$$\omega_{lj} = \left(\rho(v_l, c_j) \cdot \frac{\sum_{i=1}^n k_i}{\sum_{b=1}^n 1/\rho(v_b, c_j)} \right)^{-1}, \quad (1)$$

де ω_{lj} – міра впливу деякої категорії ознаки v_l ($l=1, \dots, \sum_{i=1}^n k_i$) на приналежність об'єкту до конкретного класу C_j ($j=1, \dots, m$), ρ – метрика у просторі кінцевої конфігурації.

Оціночні функції $W_j(x)$, що характеризують ступінь (вірогідності) приналежності деякого об'єкту $x \in X$ до класів C_j , обчислюються за формулами:

$$W_j(x) = \sum_{i=1}^n \sum_{l=1}^{k_i} \omega_{lj} \cdot \mu_{g_{li}}(x) \quad (\forall j=1, \dots, m). \quad (2)$$

Найбільш вірогідним класом для об'єкта x буде той клас C_τ , для якого одержане найбільше значення функції $W_j(x)$:

$$\tau = \arg \max_{j=1, \dots, m} W_j(x). \quad (3)$$

Співвідношеннями (1)—(3) визначається загальна модель конкретного класифікатору у вигляді алгоритму типу обчислення оцінок, яка дозволяє враховувати вплив на приналежність об'єктів до класів як якісних, так і кількісних ознак, що описують об'єкт, а також, незважаючи на представлення формули (2) у вигляді лінійної комбінації предикторів, відображає нелінійність і немонотонність зміни оцінки приналежності об'єкту до класу в залежності від описуючих змінних. Відмінність моделі, що пропонується, від наявних міститься у способі знаходження вагових коефіцієнтів змінних, які пояснюють віднесення об'єктів до різних класів. Цей спосіб базується на аналізі просторової структури взаємозв'язків між предикторними ознаками і класами, яка отримана методом множинного кореспондентського аналізу, що стає можливим за рахунок використання уявлення про класи не як про підмножини простору об'єктів, але як про ще один з показників, який складає опис об'єкту за допомогою ознак.

Метод формування композицій класифікаторів “рейтингове голосування” можна охарактеризувати як розвиток методів зваженого голосування і сумішей експертів, у яких ваги базових алгоритмів не є постійними, а залежать від об'єкту, що класифікується. Метод засновано на обчисленні рейтингів приналежності об'єкту, що класифікується, до того чи іншого класу, при розрахунку яких разом із точністю алгоритмів суміші при прогнозуванні окремих класів враховуються й похибки цих класифікаторів на інших класах. Алгоритм роботи методу схематично представлено на рис. 3.

Вважається, що є кінцева множина класифікаторів $\{T^k\}_{k=1}^N$, які вирішують задачу

дискримінації об'єктів $x \in X$ на n класів $\{C_i\}_{i=1}^n$. Для всіх T^k відома (наприклад, оцінена за навчальною вибіркою) їхня точність прогнозування класів, яка може бути задана квадратними $n \times n$ -матрицями $P^k = (p_{ij}^k)_{i,j=1}^n$, де кожен елемент p_{ij}^k розглядається як вірогідність того, що об'єкт, класифікований алгоритмом T^k , як такий, що належить до класу C_j , у дійсності належить до класу C_i . Таким чином, діагональні елементи p_{ii}^k характеризують точність алгоритмів T^k на класах C_i ($\forall k=1, \dots, N; \forall i=1, \dots, n$), а сума елементів за кожним зі стовпців матриць P^k дорівнює одиниці. З N матриць, що описано, формується єдиний трьохвимірний $N \times n \times n$ -масив P .

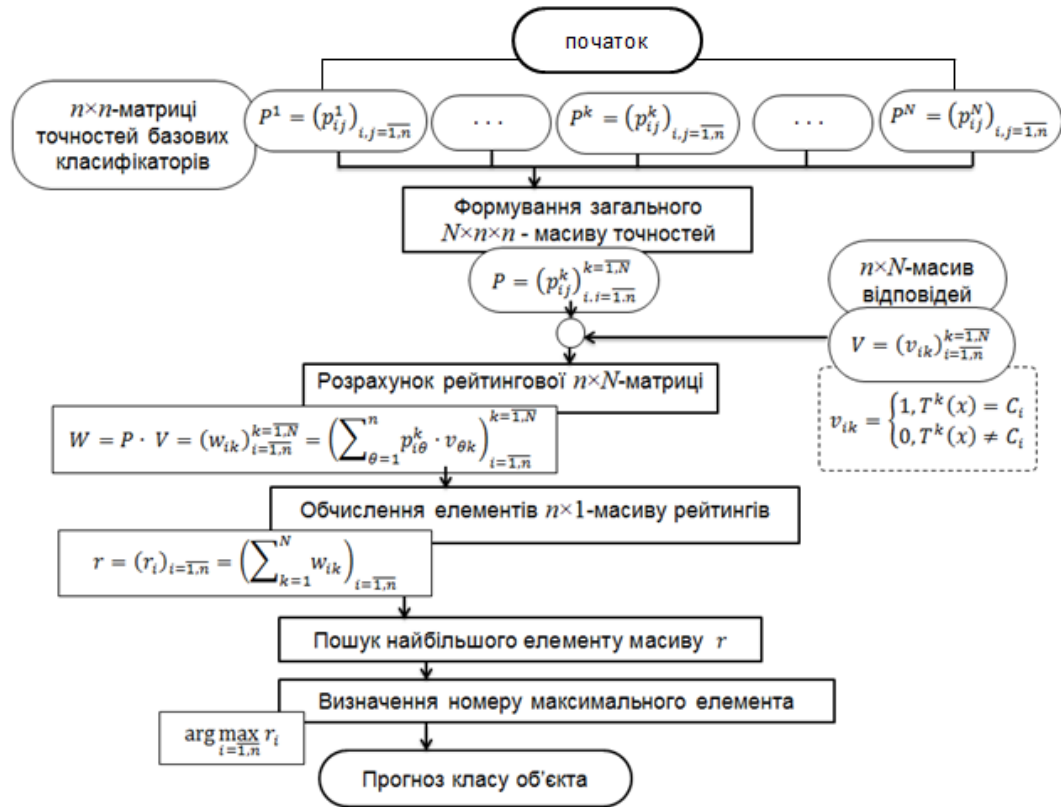


Рис. 3. Алгоритм процедури рейтингового голосування комітету класифікаторів.

Результат дії множини базових алгоритмів $\{T^k\}_{k=1}^N$ на деякому об'єкті x представляється бінарною індикаторною $n \times N$ -матрицею

$$V(x) = (v_{ij}(x))_{i=1, \dots, n}^{j=1, \dots, N}, \quad \text{де } v_{ij}(x) = \begin{cases} 1, & \text{якщо } T^j(x) = C_i \\ 0, & \text{якщо } T^j(x) \neq C_i \end{cases}. \quad (4)$$

Далі формується т.з. «рейтингова» $n \times N$ -матриця W , кожен зі стовпців якою отримується як результат множення матриці P^k на k -й стовпець матриці V :

$$W^{(k)}(x) = P^k \cdot V^{(k)}(x) \quad (5)$$

Таким чином, елементи матриці $W \forall i=1, \dots, n, \forall k=1, \dots, N$ визначаються співвідношенням:

$$w_{ik}(x) = \sum_{j=1}^n p_{ij}^k \cdot v_{jk}(x), \quad \text{що, в силу властивостей матриці } V(x), \text{ еквівалентно:}$$

$$w_{ik}(x) = \begin{cases} p_{i\theta}^k, & \text{якщо } T^k(x) = C_\theta, \text{ де } \theta \in \{1, \dots, n\} \\ 0, & \text{якщо } T^k(x) \text{ відмовляється від класифікації} \end{cases}$$

Рейтинг приналежності об'єкту x до класу C_i обчислюється як сума за i -ю строчкою матриці W :

$$r_{C_i}(x) = \sum_{k=1}^N w_{ik}(x) = \sum_{k=1}^N p_{i\theta}^k, \quad (6)$$

де $\theta \in \{0, 1, \dots, n\}$ – номер класу, до якого відносить об'єкт x алгоритм T^k (при цьому під $\theta=0$ мається на увазі відмова алгоритму від класифікації даного об'єкта).

Результатом процедури рейтингового голосування комітету класифікаторів T є клас, якому відповідає найбільший рейтинг:

$$T(x) = C_\tau, \quad \text{де } \tau = \arg \max_{i=1, \dots, n} r_{C_i}(x). \quad (7)$$

Метод створення композицій класифікаторів “рейтингове голосування по старшинству” є розвитком “рейтингового голосування” за рахунок додавання до нього принципу дії композицій із логікою старшинства. Схему його роботи відображено на рис. 4.

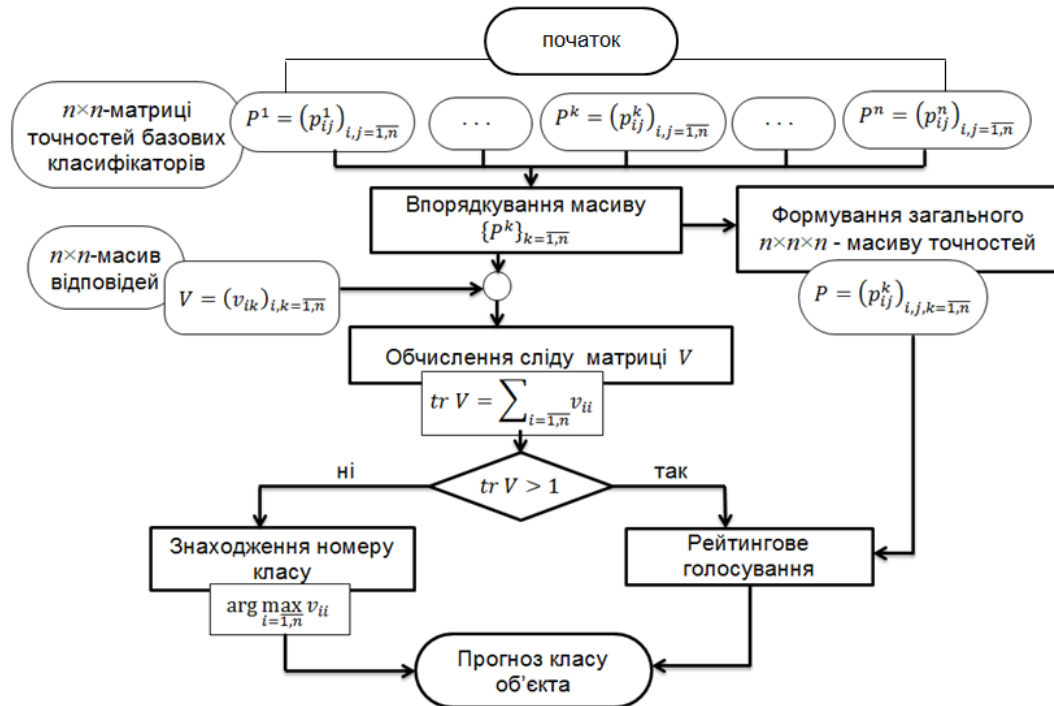


Рис. 4. Алгоритм процедури рейтингового голосування по старшинству.

Як і в комітеті з логікою старшинства, вважається, що число алгоритмів співпадає з кількістю класів, і кожен з класифікаторів найкращим чином налаштовано на розпізнавання об'єктів саме зі «свого» класу. Тобто кожен базовий алгоритм T^i має максимальну точність розпізнавання об'єктів з класу C_i , що може бути відображено співвідношенням:

$$\forall i = 1, \dots, n: p_{ii}^i \geq p_{jj}^i \quad \forall j \neq i, \quad j = 1, \dots, n. \quad (8)$$

Для досягнення властивості (8) на початковому етапі відбувається впорядкування набору базових класифікаторів у відповідності до їхніх точностей розпізнавання різних класів. Перенумерація базових класифікаторів і класів виконується таким чином, щоб для елементів загального $n \times n \times n$ -масиву точностей P виконувались властивості (8) і (9):

$$\forall i, j = 1, \dots, n: \quad i < j \Leftrightarrow p_{ii}^i > p_{jj}^j. \quad (9)$$

Останнє відповідає розташуванню базових класифікаторів у порядку зменшення точностей визначення «своїх» класів.

Результат дії набору базових класифікаторів $\{T^k\}_{k=1}^n$ на новому об'єкті x , що класифікується, представляється матрицею відповідей $V(x)$ (4), яка в даному випадку має розмір $n \times n$.

Оскільки кожен з базових класифікаторів T^k налаштовано таким чином, щоб найкращий результат видавати на класі C_k , то найбільш показовими характеристиками роботи комітету в цілому будуть значення діагональних елементів матриці V . Усі принципово різні ситуації розташування одиниць на діагоналі можна описати одним з трьох можливих значень сліду матриці

$$\text{tr}V(x) = \sum_{i=1}^n v_{ii}(x).$$

Ситуація 1: $\text{tr}V(x) = 0$. На діагоналі матриці V немає жодної одиниці, тобто жоден з базових класифікаторів не ідентифікував зразок як такий, що належить до свого класу. В цьому випадку доцільною є відмова від класифікації даного об'єкту як від нетипового (такого, що попадає у т.з. «область невпевненості», де усі алгоритми некомпетентні). Остаточна відповідь комітету: $T(x) = \emptyset$.

Ситуація 2: $\text{tr}V(x) = 1$. На діагоналі матриці V тільки одна одиниця. Класична ситуація для комітету з логікою старшинства, в якій він працює в чистому вигляді. В цьому випадку відповіддю комітету буде той клас C_τ , на якому ця одиниця з'явилась. Остаточна відповідь комітету: $T(x) = C_\tau$, де $\tau = \arg\{v_{ii} \mid v_{ii}(x) = 1\}$.

Ситуація 3: $\text{tr}V(x) > 1$. Декілька одиниць з'являються на діагоналі матриці V , коли одразу декілька базових класифікаторів відносять один об'єкт x кожен до свого класу. В такій ситуації пропонується обчислити рейтинг кожного з класів, що були спрогнозовані, у відповідності до описаної вище процедури рейтингового голосування (5)—(7) і видати остаточну

відповідь комітету: $T(x) = C_\tau$, де $\tau = \arg \max_i \sum_{j=1}^n p_{ji}^j$.

Розроблення інформаційної технології оцінювання ступеня тяжкості стану пацієнтів, яка використовує запропоновані методи побудови класифікаторів за навчальною інформацією і формування їхніх композицій, проходила поетапно у відповідності до схеми на рис. 5. На першому етапі навчальна інформація проходить попередню обробку, головною метою якої є вилучення з наявних даних елементарних правил, тобто вибір шаблонів (паттернів, найпростіших множин), які характеризують причини приналежності об'єктів до конкретних класів. Додавання до отриманого набору паттернів класів, які розглядаються як ще одна якісна ознака, що описує об'єкт, дозволяє одержати вхідну інформацію для реалізації етапу 2, на якому отримана сукупність паттернів і класів проектується в єдиний простір невеликої вимірності за допомогою методів геометричної інтерпретації структури багатовимірних даних. Отримане просторове представлення обов'язково перевіряється на якість збереження взаємозв'язків між класами і паттернами, що їх описують, які існують у вихідному просторі. При незадовільній якості одержаної проекції необхідно

повернутися до попередніх етапів і переглянути умови застосування методу (наприклад, змінити кількість вимірів простору кінцевої конфігурації), або переформулювати елементарні правила (переглянути, уточнити порогові значення, правила об'єднання чи розділення категорій, тощо). Якщо ці підходи все рівно не дозволяють досягти задовільної якості проєкції, можливо одержання декількох незалежних карт (наприклад, окремої для кожного з класів).

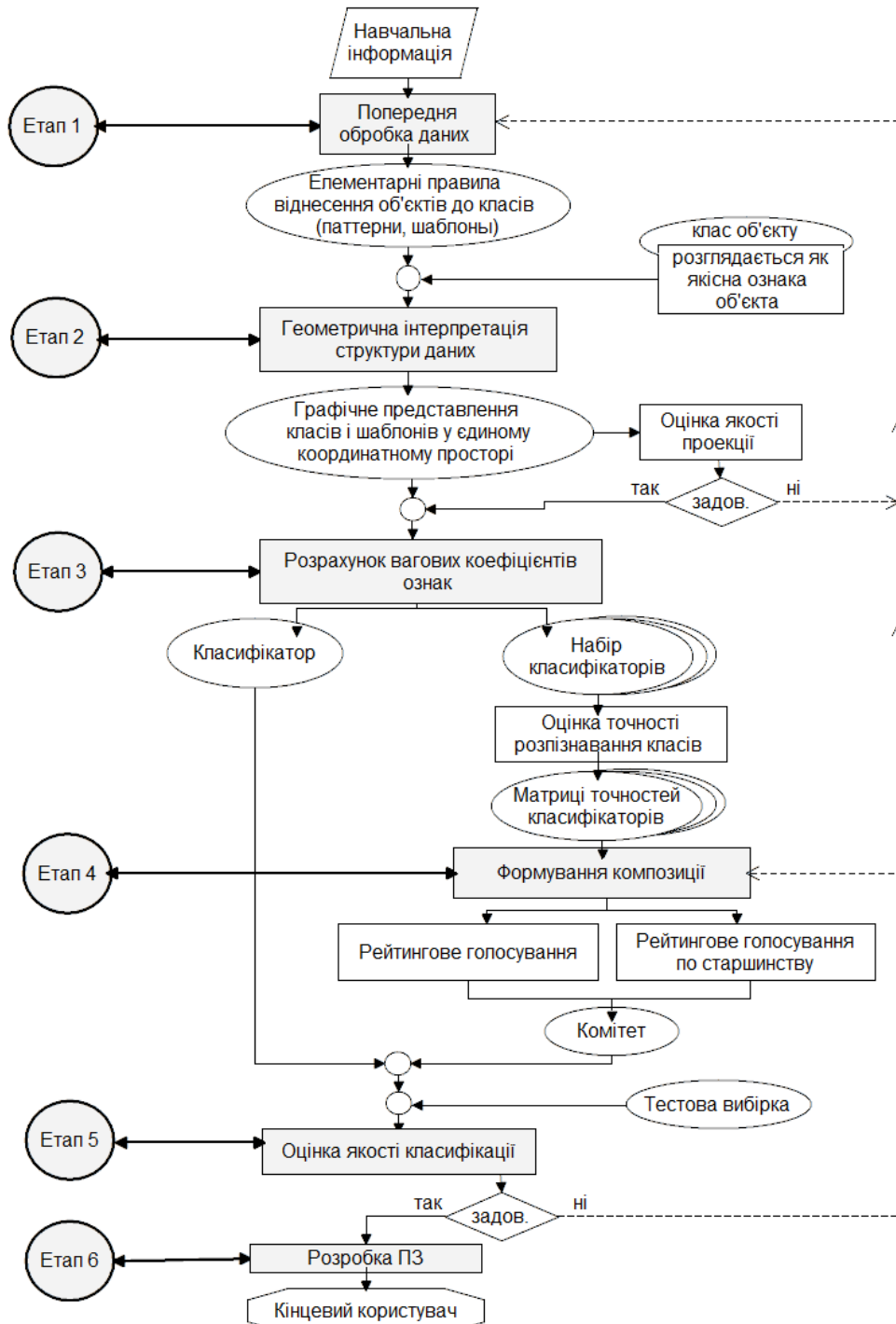


Рис. 5. Етапи розроблення інформаційної технології класифікації (оцінювання) СТС пацієнтів.

На третьому етапі розраховуються внески кожного з шаблонів у приналежність об'єкту до певного класу за формулою (1). Таким чином, на виході етапу 3 отримується або математична модель одного класифікатора у вигляді алгоритму обчислення оцінок (1)—(3), або їх набір. У

випадку декількох моделей, вони потім об'єднуються в ансамбль за допомогою одного з розроблених методів формування композицій класифікаторів (“рейтингового голосування” чи “рейтингового голосування по старшинству”).

На п'ятому етапі за допомогою тестової вибірки відбувається оцінювання якості розпізнавання класів, що надається розробленою математичною моделлю класифікатора або комітету класифікаторів. При незадовільній якості класифікації, у випадку комітету класифікаторів, можливе повернення до четвертого етапу і застосування інших методів складання композицій. У всіх випадках досягти прийнятної якості класифікації можливо шляхом повернення на етапи 2 чи 1, що дозволить отримати нові графічні представлення або формулювання елементарних шаблонів віднесення об'єктів до класів. Якщо точність класифікації, яку оцінено на п'ятому етапі за тестовою вибіркою, признано прийнятною, реалізується шостий (заклучний) етап розроблення інформаційної технології оцінювання тяжкості стану пацієнтів. На цьому етапі створюється програмне забезпечення, яке реалізує математичну модель класифікатора чи алгоритм голосування ансамблю. Розроблене в рамках запропонованої інформаційної технології програмне забезпечення надається кінцевому користувачеві для практичного застосування.

Таким чином, розроблена інформаційна технологія класифікації пацієнтів за СТС може бути реалізована у двох варіантах: з використанням тільки методу побудови класифікатора за навчальною інформацією, заснованого на геометричній інтерпретації структури багатовимірних даних (рис. 6), або для випадку декількох класифікаторів, які об'єднано в ансамбль (рис. 7).

У четвертому розділі представлені результати використання розроблених методів класифікації з навчанням і побудови композицій класифікаторів для побудови математичних моделей оцінювання СТС пацієнтів і прогнозування клінічного наслідку при травмі підшлункової залози і травматичному панкреатиті, при захворюваннях жовчовивідних проток і при інсультах.

Математичні моделі оцінки СТС і прогнозування клінічного наслідку при травматичних ушкодженнях підшлункової залози було побудовано і апробовано з використанням клінічних даних 232 пацієнтів, які проходили лікування у чотирьох клініках м. Харкова за останні декілька років. В якості потенціальних предикторів розглядалися кількісні (33) і якісні (9) показники, що описують фізіологічний стан пацієнта, параметри травматичного ушкодження, а також дані лабораторних аналізів. Оцінювання значущості впливу кожного зі згаданих вище показників на наслідок і тяжкість стану пацієнтів проводилась за допомогою різних статистичних критеріїв в залежності від шкали його вимірювання, закону розподілу і обсягу спостережень у групах. В результаті були визначені ознаки, які значущо різняться в залежності від клінічного наслідку і СТС, а також оцінені їхні порогові значення, що відповідають кожному з класів, тобто сформовані елементарні правила (паттерни), що пояснюють віднесення пацієнтів до класів.

Найбільш інформативними для прогнозування клінічного наслідку виявилися такі показники: діастолічний тиск (*BP*, мм.рт.ст.), шоківий індекс Альговера (*IS*), вік пацієнта (*Age*, років), кількість поєднаних травм (*TQ*) і обставини травми (*Circ*). Застосування множинного кореспондентського аналізу дозволило спроектувати 15 категорій досліджуваних змінних у простір 6 вимірів, зберігши при цьому 80,8% взаємозв'язків між ними, які існують у вихідному просторі. Графік проєкції ознак-предикторів і класу-наслідку у простір, що визначається першими двома власними значеннями, показано на рис. 8.

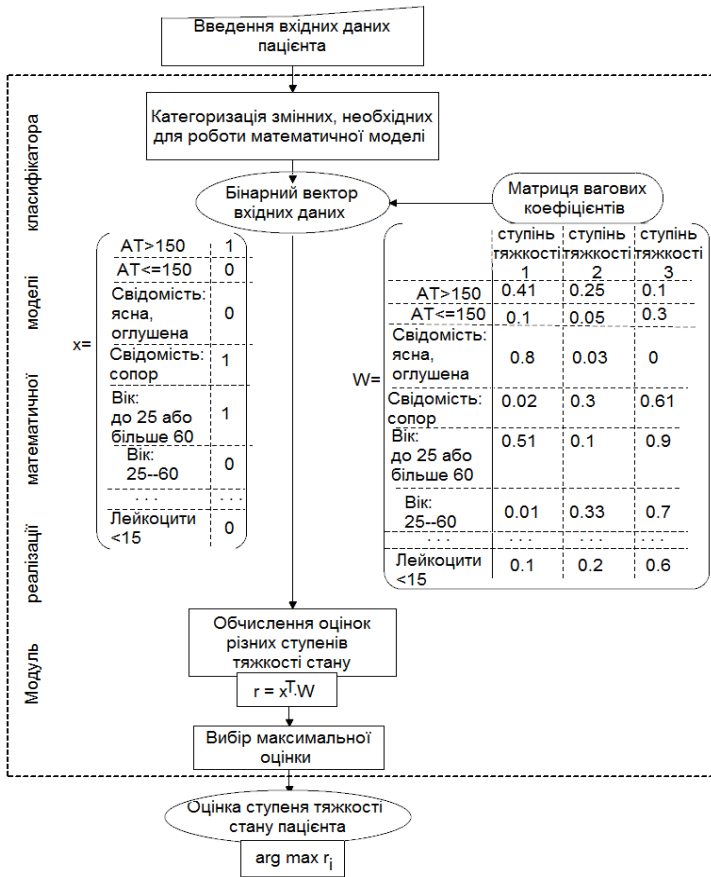


Рис. 6. Інформаційна технологія оцінювання СТС пацієнта на основі одного класифікатора.

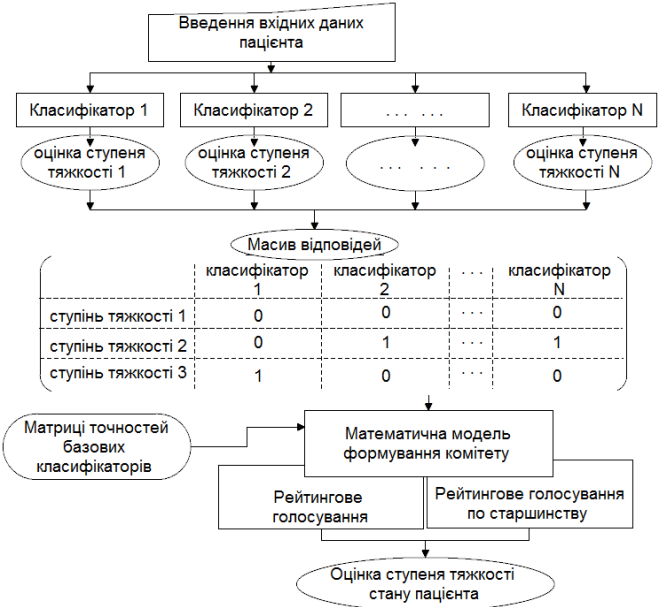


Рис. 7. Інформаційна технологія оцінювання СТС пацієнта на основі ансамблю класифікаторів.

На основі отриманого представлення у шестивимірному координатному просторі були обчислені відстані між точками-предикторами і точками-наслідками. При обчисленні відстаней використано узагальнену метрику Мінковського: $\rho(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/q}$, де показники ступеня було обрано в результаті чисельних експериментів для забезпечення найбільшої точності: $p = 2$, $q = 1$, $\rho(x, y) = \sum_{i=1}^6 |x_i - y_i|^2$.

Таким чином, одержано математичну модель прогнозування летального наслідку при травмі підшлункової залози, яку формально можна представити у вигляді лінійної комбінації 11-ти дихотомічних змінних:

$$I_{leth} = \sum_{i=1}^{11} \alpha_i \cdot x_i = 0,015 \cdot x_1 + 0,231 \cdot x_2 - 0,683 \cdot x_3 - 0,044 \cdot x_4 + 0,138 \cdot x_5 + 0,035 \cdot x_6 + 0,035 \cdot x_7 + 0,073 \cdot x_8 + 0,026 \cdot x_9 - 0,024 \cdot x_{10} - 0,043 \cdot x_{11} + 0,120$$

де $x_1 = \begin{cases} 1, & BP < 40 \\ 0, & BP \geq 40 \end{cases}$, $x_2 = \begin{cases} 1, & 40 \leq BP \leq 64 \\ 0, & BP > 64 \vee BP < 40 \end{cases}$, $x_3 = \begin{cases} 1, & BP \geq 65 \\ 0, & BP < 65 \end{cases}$, $x_4 = \begin{cases} 1, & IS \leq 0,8 \\ 0, & IS > 0,8 \end{cases}$,
 $x_5 = \begin{cases} 1, & 0,8 < IS \leq 1,25 \\ 0, & IS \leq 0,8 \vee IS > 1,25 \end{cases}$, $x_6 = \begin{cases} 1, & IS > 1,25 \\ 0, & IS \leq 1,25 \end{cases}$, $x_7 = \begin{cases} 1, & Age \leq 31 \\ 0, & Age > 31 \end{cases}$, $x_8 = \begin{cases} 1, & 32 \leq Age \leq 57 \\ 0, & Age < 32 \vee Age > 57 \end{cases}$,

$$x_9 = \begin{cases} 1, & \text{Age} > 57 \\ 0, & \text{Age} \leq 57 \end{cases}; \quad x_{10} = \begin{cases} 1, & TQ > 2 \\ 0, & TQ \leq 2 \end{cases}; \quad x_{11} = \begin{cases} 1, & \text{Circ} = \text{падіння з висоти} \\ 0, & \text{Circ} = \text{інше} \end{cases}.$$

У відповідності до побудованої математичної моделі при обчисленому значенні $I_{leth} > 0$ треба прогнозувати летальний наслідок, при $I_{leth} < 0$ – більшу вірогідність сприятливого наслідку.

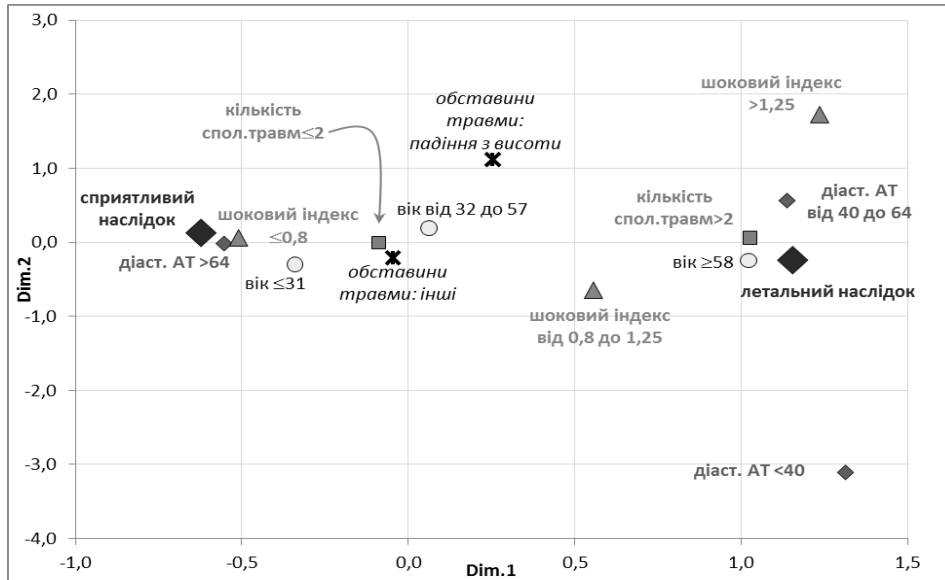


Рис. 8. Проекція ознак-предикторів клінічного наслідку у простір, що визначається власними значеннями 1 і 2.

В результаті застосування побудованої моделі до вибірки з 216 пацієнтів із травматичними пошкодженнями підшлункової залози і травматичним панкреатитом вірно було спрогнозовано 122 сприятливих наслідки з 142, що відповідає 85,9% специфічності класифікатора; і 65 з 74 летальних наслідків, що відповідає 87,8% чутливості. Таким чином, середня точність прогнозу в побудованій моделі дорівнює 86,9%.

При розробленні моделі класифікації пацієнтів на три СТС (“≤3” – стан легкої і середньої тяжкості, “4” – тяжкий, “5” – критичний стан) за допомогою запропонованого методу класифікації з навчанням на основі геометричної інтерпретації структури багатовимірних даних були побудовані три однокласових класифікатори, які дозволяють відділити кожен зі ступенів тяжкості стану від інших. Для формування загальної моделі оцінки СТС пацієнтів з травматичними пошкодженнями підшлункової залози і травматичним панкреатитом однокласові класифікатори для визначення трьох ступенів тяжкості було об’єднано в комітет за допомогою методу рейтингового голосування по старшинству. Виходячи з прогностичної точності (user’s accuracy) однокласових класифікаторів I_3 , I_4 , I_5 , яка була оцінена на тестовій вибірці, було визначено правила прогнозування СТС у випадку віднесення пацієнта до більш ніж одного класу, що дозволило сформувати модель оцінювання СТС пацієнтів, яка працює за схемою на рис. 9.

Розпізнавання СТС пацієнта із травмою підшлункової залози і травматичним панкреатитом за допомогою побудованого комітету класифікаторів дозволяє визначати легкі (“≤3”) ступені тяжкості з точністю 88,24%, тяжкі (ступені тяжкості 4) стани – з точністю 83,33%, і критичні (ступені тяжкості 5) стани – з точністю 89,71%. Таким чином, в середньому точність класифікації складає 87,09%.

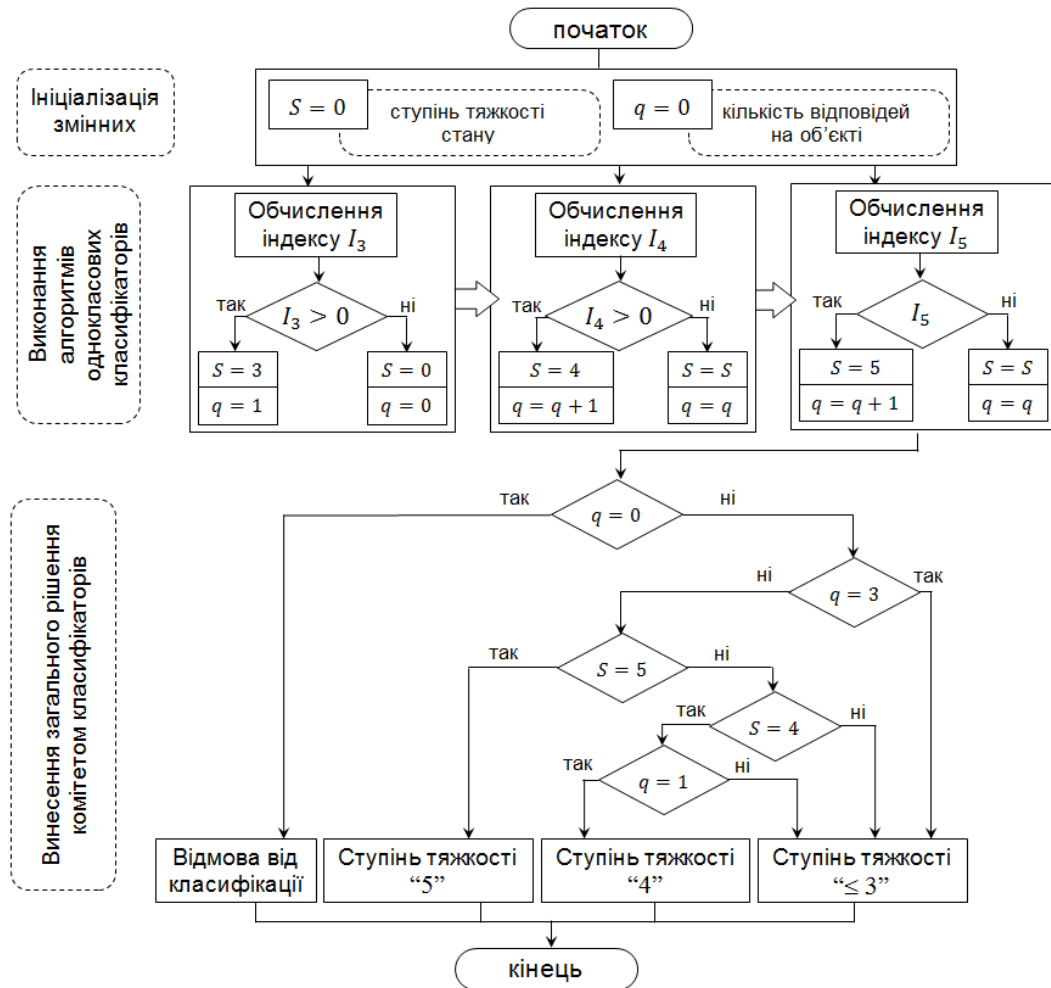


Рис. 9. Алгоритм роботи комітету класифікаторів для оцінки СТС при травмі підшлункової залози.

Побудовані математичні моделі та алгоритми оцінювання СТС і прогнозування клінічного наслідку травми підшлункової залози і травматичного панкреатиту виступили в якості методичного забезпечення для розроблення програмного засобу, орієнтованого на медичний персонал, що дозволяє здійснити підтримку прийняття рішень лікаря стосовно пацієнтів із даним захворюванням. Програму написано на мові програмування Visual Basic for Applications (VBA) для роботи з базами даних форматів *.mdb і *.accdb СКБД Microsoft Office Access 2003–2012. Її основними функціями є прогнозування клінічного наслідку травматичного панкреатиту і травми підшлункової залози на основі побудованої математичної моделі класифікації; оцінювання СТС пацієнтів за допомогою комітету класифікаторів, який здійснює рейтингове голосування по старшинству та об'єднує три математичні моделі класифікації; введення і збереження в базі даних відомостей про нових пацієнтів, а також редагування даних про наявних пацієнтів; додатково можливе групування і вибір даних з БД за різними критеріями із переглядом звітів за допомогою вбудованих засобів СКБД MS Office Access 2003–2012.

Математичну модель диференційної діагностики захворювань жовчовивідних проток побудовано з використанням клінічних даних 90 пацієнтів, які страждають на захворювання жовчовивідних проток різного ступеня тяжкості. В залежності від форми захворювання всі пацієнти були розділені на три класи: 1 – пацієнти з найбільш легкою формою – механічною жовтяницею (24 пацієнти); 2 – пацієнти з діагнозом гострий холангіт (46 пацієнтів); 3 – пацієнти з

найбільш тяжким діагнозом – біліарний сепсис (20 пацієнтів). В якості описуючих ознак використовувались дані лабораторних аналізів при надходженні в стаціонар і результати оцінки їх тяжкості стану за бальними шкалами SOFA (Sepsis Oriented Failure Assessment) і SIRS (Systemic Inflammation Response Syndrome) (загалом 37 інтервальних і порядкових змінних). Для пошуку кількісних змінних, які можуть обумовлювати різницю між класами, використовували дисперсійний аналіз та його непараметричний аналог – аналіз Краскала–Уолліса. Для змінних, які значущо відрізняються в залежності від класу, з метою деталізації цієї різниці, проводились попарні порівняння між класами. Для виявлення різниць між парами класів застосовувався непараметричний тест незалежних вибірок Манна–Уїтні з поправкою Бонферроні на множинність. Для дослідження впливу порядкових ознак на форму захворювання (клас) проводився аналіз таблиць спряженості. Значущість впливу підтверджувалась за допомогою критерію χ^2 Пірсона. При виділенні інформативних інтервалів значень змінних, які пояснюють наявність в пацієнта тієї чи іншої форми захворювання (його приналежність к класу 1, 2 чи 3), визначалися пороги показників, а потім аналізувалися двохходові таблиці спряженості форми захворювання (класу) з кожною категоризованою ознакою. Усі обчислення проводились при довірчій ймовірності 95%. На основі проведених досліджень отримано статистично обґрунтовані розділення кількісних показників на інтервали значень (елементарні правила, що описують приналежність об'єктів до класів), які потім разом із класами об'єктів за допомогою множинного кореспондентського аналізу були представлені в єдиному координатному просторі. Для збереження 94,5% взаємозв'язків між 16 досліджуваними категоріями знадобилося 7 вимірів простору кінцевої конфігурації, в якому з використанням метрики Евкліда було обчислено відстані між класами, що прогножуються, і паттернами, що їх описують, на основі яких за формулою (1) одержано вагові коефіцієнти функцій класифікації:

Для класу 1 (механічна жовтяниця):

$$S_1 = -1,29 \cdot NCB - 2,43 \cdot ALT + 3,40 \cdot PI + 4,19 \cdot SOFA + 24,34 \cdot SIRS_0 + 5,17 \cdot SIRS_1 + 3,35 \cdot SIRS_2 - 0,98 \cdot SB + 32,12.$$

Для класу 2 (гострий холангіт):

$$S_2 = -5,39 \cdot NCB + 4,73 \cdot ALT + 3,74 \cdot PI + 7,20 \cdot SOFA + 5,32 \cdot SIRS_0 + 20,50 \cdot SIRS_1 + 2,58 \cdot SIRS_2 + 0,74 \cdot SB + 30,30.$$

Для класу 3 (біліарний сепсис):

$$S_3 = 0,96 \cdot NCB + 0,50 \cdot ALT - 0,30 \cdot PI + 2,38 \cdot SOFA + 6,25 \cdot SIRS_0 + 6,64 \cdot SIRS_1 + 16,80 \cdot SIRS_2 + 1,03 \cdot SB + 32,87,$$

$$\text{де } NCB = \begin{cases} 1, & \text{якщо непряма білірубін} > 40 \text{ мкмоль/л} \\ 0, & \text{якщо непряма білірубін} \leq 40 \text{ мкмоль/л} \end{cases}, \quad ALT = \begin{cases} 1, & \text{якщо АЛТ} > 1,5 \text{ мкмоль/мл} \cdot \text{ч} \\ 0, & \text{якщо АЛТ} \leq 1,5 \text{ мкмоль/мл} \cdot \text{ч} \end{cases},$$

$$PI = \begin{cases} 1, & \text{якщо ПІ} \geq 83\% \\ 0, & \text{якщо ПІ} < 83\% \end{cases}, \quad SOFA = \begin{cases} 1, & \text{якщо SOFA} \leq 3 \\ 0, & \text{якщо SOFA} > 3 \end{cases}, \quad SB = \begin{cases} 1, & \text{якщо п/я} \geq 5\% \\ 0, & \text{якщо п/я} < 5\% \end{cases},$$

$$SIRS_0 = \begin{cases} 1, & \text{якщо SIRS} = 0 \\ 0, & \text{якщо SIRS} \neq 0 \end{cases}, \quad SIRS_1 = \begin{cases} 1, & \text{якщо SIRS} = 1 \\ 0, & \text{якщо SIRS} \neq 1 \end{cases}, \quad SIRS_2 = \begin{cases} 1, & \text{якщо SIRS} = 2 \\ 0, & \text{якщо SIRS} \neq 2 \end{cases}.$$

У відповідності до співвідношення (3) новий пацієнт повинен бути віднесений до класу з найбільшим значенням оціночної функції.

На навчальній виборці в першому класі правильно були класифіковані усі 24 пацієнта (100%), в другому – 33 (72%) і в третьому – 15 (75%). Таким чином, правильно було класифіковано 72 з 90 пацієнтів, що дозволяє стверджувати про 80% загальної точності методу.

Математична модель визначення клінічного наслідку інсульту була побудована на основі клінічних даних 1091 пацієнта з ішемічними і геморагічними інсультами, госпіталізованих протягом 2008 року в ГКЛ № 7 м. Харкова. У досліджуваній виборці містилися дані про 870 сприятливих наслідків інсульту і про 221 фатальний наслідок. Вхідні показники для побудови класифікатора обиралися з сукупності симптомів інсульту, які оцінювалися лікарем при надходженні пацієнта в стаціонар, а також з набору факторів ризику інсульту, відомих з анамнезу. Усі показники були виміряні в номінальній або, максимум, в порядковій шкалі, тому основним інструментом їх попереднього аналізу були методи аналізу таблиць спряженості ознак. В якості додаткових параметрів використовувалися бальні оцінки самопочуття пацієнтів, отримані за шкалами GCS (Glasgow Coma Scale, ШКГ) і NIHSS (National Institutes of Health Stroke Scale).

Застосування запропонованого методу класифікації з навчанням на основі геометричної інтерпретації структури багатовимірних даних дозволило отримати формулу для обчислення прогностичного індексу наслідку інсульту (I):

$$I = 0,095 \cdot x_1 - 0,332 \cdot x_2 - 0,003 \cdot x_3 + 0,053 \cdot x_4 + 0,048 \cdot x_5 + 0,010 \cdot x_6 + 0,132 \cdot x_7 + 0,092 \cdot x_8 + 0,107 \cdot x_9 + 0,039 \cdot x_{10} - 0,116$$

де x_i ($i = 1, \dots, 10$) – дихотомічні змінні, що приймають значення 0 або 1:

$$x_1 = \begin{cases} 1, & \text{при інсульті ішемічного типу} \\ 0, & \text{при інсульті геморагічного типу} \end{cases}, \quad x_2 = \begin{cases} 1, & \text{при балі ШКГ} \leq 13 \\ 0, & \text{при балі ШКГ} > 13 \end{cases}$$

$$x_3 = \begin{cases} 1, & \text{при балі ШКГ} = 14 \\ 0, & \text{при інших балах ШКГ} \end{cases}, \quad x_4 = \begin{cases} 1, & \text{при балі ШКГ} = 15 \\ 0, & \text{при інших балах ШКГ} \end{cases}$$

$$x_5 = \begin{cases} 1, & \text{при балі NISS} \leq 9 \\ 0, & \text{при балі NISS} \geq 10 \end{cases}, \quad x_6 = \begin{cases} 1, & \text{за відсутності запаморочення} \\ 0, & \text{при запамороченні} \end{cases}$$

$$x_7 = \begin{cases} 1, & \text{за відсутності парезу VII пари} \\ 0, & \text{при наявності парезу VII пари} \end{cases}, \quad x_8 = \begin{cases} 1, & \text{при нормальній ротації язика} \\ 0, & \text{при порушеній ротації язика} \end{cases}$$

$$x_9 = \begin{cases} 1, & \text{за відсутності атаксії} \\ 0, & \text{при наявності атаксії} \end{cases}, \quad x_{10} = \begin{cases} 1, & \text{за відсутності аритмії на ЕКГ} \\ 0, & \text{при наявності аритмії на ЕКГ} \end{cases}$$

Обчислене таким чином значення індексу I теоретично може змінюватися в границях від -1 до $+1$ (практично від $-0,46$ до $+0,46$), причому, додатні значення відповідають сприятливому прогнозу, від'ємні – прогнозу летального наслідку інсульту.

Перевірка якості прогнозування наслідку за допомогою запропонованої моделі показала, що загальна точність методу складає 91,93%, точність покриття класу сприятливих наслідків (специфічність) дорівнює 95,98%, точність покриття класу летальних наслідків (чутливість) дорівнює 76,02%.

Для порівняння точності моделей, побудованих на основі запропонованих методів класифікації з навчанням і побудови композицій класифікаторів, для прогнозування клінічного наслідку травми підшлункової залози і травматичного панкреатиту були побудовані альтернативні моделі за допомогою методів логіт-регресії і дерев класифікації; для оцінювання СТС пацієнтів з

травмами підшлункової залози і травматичним панкреатитом, а також для диференційної діагностики захворювань жовчовивідних проток були побудовані альтернативні моделі за допомогою лінійного дискримінантного аналізу Фішера; для прогнозування клінічного наслідку інсульту альтернативну модель було побудовано за допомогою метода дерев рішень. Порівняльний аналіз показав перевагу запропонованої технології перед класичними методами статистичного моделювання, що виражається у підвищенні загальної точності класифікації на $4,4 \div 8,9\%$, специфічності – на $4,2 \div 6\%$, чутливості – на $7,6 \div 24,8\%$.

ВИСНОВКИ

У дисертації запропоновано нове рішення актуальної науково-прикладної задачі – розроблення моделей і методів оцінювання ступеня тяжкості стану пацієнтів для підтримки прийняття рішень лікаря та їх реалізації в інформаційній технології для підвищення точності визначення ступеня тяжкості стану пацієнтів і прогнозування клінічного наслідку за рахунок використання моделей, побудованих на основі запропонованих методів класифікації з навчанням і формування композицій класифікаторів.

1. Аналіз наявних методів і моделей оцінювання тяжкості стану пацієнтів дозволив виділити з методів побудови моделей класифікації з навчанням методи KDD (Knowledge Discovery in Databases), які не накладають жорстких вимог на вид моделі та на властивості вхідних даних, у якості найбільш перспективних для медичних застосувань, зокрема клас методів графічної інтерпретації та візуалізації багатовимірних даних, перспективи та результати практичного застосування яких ще недостатньо вивчені.

2. Запропонований метод побудови класифікаторів за навчальною інформацією, заснований на геометричній інтерпретації структури багатовимірних даних і уявленні про клас не як про підмножину простору об'єктів, але як про одну зі складових його опису за допомогою ознак, забезпечує сумісне використання якісних і кількісних незалежних змінних і дає можливість створювати алгоритми типу обчислення оцінок для класифікації пацієнтів в залежності від їхньої тяжкості стану, а також отримувати розв'язки інших задач класифікації з навчанням, що мають підвищену точність.

3. Розроблений метод кількісної оцінки сили впливу описуючих ознак на приналежність об'єктів до класів, який базується на метричному підході до аналізу просторової структури взаємозв'язків між предикторними ознаками та класами, дозволяє враховувати нелінійність і немонотонність поведінки предикторних змінних при переході від класу до класу, що сприяє підвищенню якості класифікації.

4. Запропоновані методи формування композицій класифікаторів «рейтингове голосування» та «рейтингове голосування по старшинству», які є розвитком процедур зваженого голосування, сумішей експертів та голосування по старшинству та засновані на поєднанні принципів спеціалізації та зваженого голосування, за рахунок сумісного врахування не тільки точності базових алгоритмів при прогнозуванні окремих класів, а також і їхніх похибок на інших класах, дозволяють підвищити якість класифікації та зменшити кількість відмов від класифікації у порівнянні зі стандартними композиціями з логікою старшинства чи більшості у випадках одночасного використання декількох класифікаторів.

5. Застосування побудованих моделей і алгоритмів оцінювання СТС пацієнтів, які базуються

на використанні запропонованих методу побудови класифікаторів на основі аналізу просторового представлення ознакових змінних і методів формування композицій класифікаторів на основі поєднання принципів спеціалізації та зваженого голосування, а також інформаційної технології, яка реалізує їх у автоматизованій системі, дозволило підвищити точність оцінки СТС пацієнтів при травматичних пошкодженнях підшлункової залози та травматичному панкреатиті, при захворюваннях жовчовивідних проток та при інсультах.

6. Порівняльний аналіз виявив, що розроблені на основі запропонованого методу побудови класифікаторів моделі мають більшу точність у порівнянні з моделями, які побудовані на основі класичних методів статистичного моделювання (аналіз дискримінантних функцій, логістична регресія, метод дерев рішень). За показниками загальної точності класифікації перевищення складає $4,4 \div 8,9\%$, за специфічністю – $4,2 \div 6\%$, за чутливістю – $7,6 \div 24,8\%$.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Краснобаев В.А. Метод экспресс-диагностики оценки тяжести состояния пострадавших с тупыми травмами поджелудочной железы / В.А. Краснобаев, М.Н. Нессонова // Системы обработки информации. – 2008. – Вып. 7(74). – С. 180-185.
2. Краснобаев В.А. Оценка значущости влияния поединных травм на ступень тяжести stanu пациентов із пошкодженнями підшлункової залози / В.А. Краснобаев, М.М. Нессонова // Системы озброєння і військова техніка. – 2008. – Вып. 4(16). – С. 167-170.
3. Краснобаев В.А. Розробка комбінованої моделі дискримінантних функцій для оцінки тяжкості стану пацієнтів при травматичному панкреатиті / В.А. Краснобаев, М.М. Нессонова // Системы управління, навігації та зв'язку. – 2009. – Вып. 1(9). – С. 98-104.
4. Арсеньев А.В. Методы срочной (экспресс) оценки тяжести состояния пациентов с травматическими повреждениями поджелудочной железы. Часть I. Метод экспресс-оценки тяжести состояния при травмах поджелудочной железы, основанный на двух моделях дискримінантних функцій / А.В. Арсеньев, С.Н. Тесленко, М.Н. Нессонова // Радиоэлектроника и информатика. – 2010. – № 1. – С. 101-107.
5. Арсеньев А.В. Методы срочной (экспресс) оценки тяжести состояния пациентов с травматическими повреждениями поджелудочной железы. Часть II. Упрощенный балльный экспресс-метод оценки тяжести состояния при травмах поджелудочной железы / А.В. Арсеньев, С.Н. Тесленко, М.Н. Нессонова // Радиоэлектроника и информатика. – 2010. – № 2. – С. 77-80.
6. Арсеньев А.В. Методы срочной (экспресс) оценки тяжести состояния пациентов с травматическими повреждениями поджелудочной железы. Часть III. Метод экспресс-оценки тяжести состояния при травмах поджелудочной железы, основанный на применении многомерного шкалирования / А.В. Арсеньев, С.Н. Тесленко, М.Н. Нессонова // Радиоэлектроника и информатика. – 2010. – № 3. – С. 51-56.
7. Кочина М.Л. Результаты использования информационной технологии оценки тяжести состояния пациентов и прогноза исхода при травме поджелудочной железы / М.Л. Кочина, М.Н. Нессонова // Клиническая информатика и телемедицина. – 2013. – № 9, Вып. 10. – С. 102-107.
8. Кочина М.Л. Методы оценки степени тяжести состояния пациентов / М.Л. Кочина,

М.Н. Нессонова // Кибернетика и вычислительная техника. – 2014. – Вып. 175. – С. 74-87.

9. Пат. 39948 Україна МПК А61В 5/00, G09В 23/00. Спосіб оцінки стану підшлункової залози при механічних пошкодженнях / Криворучко І.А., Арсеньев О.В., Тесленко С.М., Гончарова Н.М., Сивожелізов А.В., Вовк В.А., Грінченко С.В., Близнюк В.В., Нессонова М.М. – № 200809256; – заявл. 15.07.2008; опубл. 25.03.2009, бюл. № 6/2009.

10. Нессонова М.Н. Метод построения классификаторов на основании геометрической интерпретации структуры данных / Нессонова М.Н. // Актуальні питання фармацевтичної і медичної науки та практики – 2014. – № 1 (14). – С. 101-103.

11. Нессонова М.Н. Информационная технология оценки степени тяжести состояния пациентов / Нессонова М.Н. // Scientific Journal «ScienceRise». – 2014. – № 2(2). – С. 37-43.

12. Арсеньев А.В. Построение системы оценки тяжести состояния пациентов при травматическом панкреатите. / Арсеньев А.В., Тесленко С.Н., Нессонова М.Н. // «Актуальные проблемы биомедицины»: конф. 3-его Международного форума «Прикладная радиоэлектроника. Состояние и перспективы развития» МРФ-2008, 22–24 октября 2008 г., Харьков : сб. научн. трудов ; т. IV. – Х., АНПРЭ, ХНУРЭ, 2008. – С.254-256.

13. Арсеньев А.В. Математическая модель оценки тяжести состояния при травматическом панкреатите на основе комбинированного применения методов многомерного статистического анализа / А.В. Арсеньев, М.Н. Нессонова // «Актуальні питання дистанційної освіти та телемедицини 2009»: Всеукраїнська науково-практична відео конференція ; 16–17 квітня 2009 р., Запоріжжя – Київ : тези доповідей. – Запоріжжя, ЗДМУ, 2009. – С. 104-107.

14. Арсеньев А.В. Компьютерная реализация методов оценки тяжести состояния пациентов с травмами поджелудочной железы / А.В. Арсеньев, С.Н. Тесленко, М.Н. Нессонова // «Медична та біологічна інформатика і кібернетика: віхи розвитку»: конференція з міжнародною участю ; 20–23 квітня 2011 р., Київ : збірник праць. – К., НМПО ім. П.Л. Шупика, 2011. – С. 88-89.

15. Нессонова М.Н. Метод рейтингового голосования комитета алгоритмов в задачах классификации с учителем / Нессонова М.Н. // «Актуальні питання дистанційної освіти і телемедицини' 2012»: всеукр. наук.-метод. відеоконф. с міжнар. участю; 17–19 жовтня 2012 р., Запоріжжя : Запорожский медицинский журнал. – 2013. – № 1(76). – С. 101-102.

16. Кочина М.Л. Информационная технология оценки тяжести состояния пациента / М.Л. Кочина, М.Н. Нессонова // «Информационные технологии в кардиологии»: международная конф.; 11–12 апреля 2013 г., Харьков : Клиническая информатика и телемедицина. – 2013. – № 9, Вып. 10. – С. 180-181.

17. Nessonova M.N. Metric Approach Based Classifier / Nessonova M.N. // “Actual Questions of Development of New Medicines” («Актуальні питання створення нових ЛЗ»): всеукр. наук.-практ. конф. студентів та молодих вчених; 25–26 квітня 2013 р., Харків : тези доповідей. – Х., НФаУ, 2013. – С. 269.

18. Nessonova M.N. The Method to Develop Classifiers Based On Geometrical Interpretation of Data Structure / Nessonova M.N. // «Актуальні питання дистанційної освіти та телемедицини 2013»: всеукр. наук.-метод. відеоконф. с міжнар. участю; 10–11 жовтня 2013 р., Запоріжжя – Київ : Актуальні питання фармацевтичної і медичної науки та практики. – 2013. – № 3 (13), додаток. – С. 63-64.

19. Кочина М.Л. Оценка тяжести состояния пациента и прогноз исхода методами

классификации с обучением на основе метрического подхода к геометрической интерпретации структуры данных / М.Л. Кочина, М.Н. Нессонова // «Информационные технологии в неврологии, психиатрии, эпилептологии и медицинской статистике»: научно-практ. конф. с международным участием; 17–18 октября 2013, Киев: Клиническая информатика и телемедицина. – 2014. – Т. 10, Вып. 11. – С. 157-158.

20. Нессонова М.Н. Особенности применения методов математической статистики для доказательной медицины / М.Н. Нессонова, Ю.М. Пенкин // «Актуальні питання дистанційної освіти і телемедицини»: всеукр. наук.-метод. відеоконф. с міжнар. участю; 16–17 жовтня 2014 р., Запоріжжя: тези доповідей. – Запоріжжя, ЗДМУ, 2014. – С. 126-127.

АНОТАЦІЯ

Нессонова М.М. Моделі і методи оцінки ступеня тяжкості стану пацієнтів для підтримки прийняття рішень лікаря. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.09 – медична і біологічна інформатика та кібернетика. – Міжнародний науково-навчальний центр інформаційних технологій і систем НАН України і МОН України, Київ, 2015.

У дисертації запропоновано нове рішення актуальної науково-прикладної задачі – розроблення моделей і методів оцінювання ступеня тяжкості стану пацієнтів для підтримки прийняття рішень лікаря.

Запропоновано метод класифікації з навчанням, який базується на аналізі просторового представлення взаємозв'язків між класами та елементарними правилами, що їх описують, яке отримується за допомогою методів геометричної інтерпретації структури багатовимірних даних. Отримали подальший розвиток методи формування композицій класифікаторів за рахунок комбінування принципів спеціалізації та зваженого голосування. Запропоновані методи побудови класифікаторів за навчальною інформацією та їхніх композицій реалізовані в інформаційній технології, що дозволило побудувати математичні моделі для оцінювання ступеня тяжкості стану та прогнозування клінічного наслідку пацієнтів із травмами підшлункової залози і травматичним панкреатитом, захворюваннями жовчовивідних проток, з інсультами, які мають підвищену точність у порівнянні з моделями, побудованими на основі класичних методів статистичного моделювання.

Ключові слова: класифікація з навчанням (зі вчителем), композиції класифікаторів, моделі оцінювання ступеня тяжкості стану пацієнта, інформаційна технологія.

АННОТАЦИЯ

Нессонова М.Н. Модели и методы оценки степени тяжести состояния пациентов для поддержки принятия решений врача. – На правах рукописи.

Диссертация на соискание учёной степени кандидата технических наук по специальности 05.13.09 – медицинская и биологическая информатика и кибернетика. – Международный научно-учебный центр информационных технологий и систем НАН Украины и МОН Украины, Киев, 2015.

В диссертации предложено новое решение актуальной научно-прикладной задачи – разработки моделей и методов оценки степени тяжести состояния пациентов для поддержки

принятия решений врача и их реализации в информационной технологии с целью повышения точности определения степени тяжести состояния пациентов и прогнозирования клинического исхода, что достигается за счёт использования моделей, построенных на основе разработанных методов классификации с обучением и формирования композиций классификаторов.

Проведенный анализ существующих методов и моделей оценки тяжести состояния пациентов показал целесообразность решения задачи оценки степени тяжести состояния как задачи классификации с обучением. Из существующих методов построения моделей классификации с обучением, как наиболее перспективные в медицинских приложениях, выделены методы KDD (Knowledge Discovery in Databases), которые жёстко не регламентируют вид модели и свойства исходных данных. Среди этих методов выявлен класс методов графической интерпретации и визуализации многомерных данных, перспективы и результаты практического применения которых ещё недостаточно изучены.

Предложенный метод построения классификаторов по обучающей информации, основанный на геометрической интерпретации структуры многомерных данных, базируется на анализе структуры подмножеств признаков в пространстве объектов с использованием дополнительного признака в виде вектора меток классов, что отличает его от традиционного подхода, использующего представление о классах как о подмножествах объектов в пространстве признаков. Данный метод построения моделей классификации обеспечивает совместное использование качественных и количественных независимых переменных и даёт возможность создавать алгоритмы вычисления оценок для классификации пациентов в зависимости от тяжести их состояния, а также получать решения других задач классификации с обучением, имеющие повышенную точность. Разработанный метод количественной оценки силы влияния описывающих признаков на принадлежность объектов к классам, основанный на метрическом подходе к анализу пространственной структуры взаимосвязей между предикторными признаками и классами, позволяет учитывать нелинейность и немонотонность поведения предикторных переменных при переходе от класса к классу, что способствует повышению качества классификации.

Предложенные методы формирования композиций классификаторов “рейтинговое голосование” и “рейтинговое голосование по старшинству” являются развитием процедур взвешенного голосования, смесей экспертов и голосования по старшинству. Они основаны на синтезе эвристик, лежащих в основе принципов специализации и взвешенного голосования. В методе “рейтингового голосования” при формировании общего решения комитета классификаторов предлагается учитывать не только точности базовых алгоритмов при прогнозировании отдельных классов, но и их ошибки на других классах. Это позволяет повысить качество классификации и уменьшить количество отказов от классификации по сравнению со стандартными композициями с логикой старшинства или большинства в случаях одновременного использования нескольких моделей классификации.

Применение предложенного метода построения классификаторов, который базируется на анализе пространственного представления признаков переменных, и методов формирования композиций классификаторов, основанных на совмещении принципов специализации и взвешенного голосования, позволило разработать модели и алгоритмы оценки степени тяжести состояния пациентов и прогнозирования клинического исхода при травматических повреждениях поджелудочной железы и травматическом панкреатите, при заболеваниях желчевыводящих

протоколов и при инсультах. Одним из преимуществ разработанных моделей классификации пациентов по степени тяжести состояния и клиническому исходу является немаловажное для практического применения свойство интерпретируемости, т.к. пространственные визуализации взаимосвязей между прогнозируемыми классами и описывающими их паттернами, на основе которых строятся данные модели, позволяют получить наглядное объяснение, почему данный прогноз был выдан подобно тому, как это происходит в экспертных системах. Построенные математические модели и алгоритмы оценки степени тяжести состояния и прогнозирования клинического исхода травмы поджелудочной железы и травматического панкреатита явились методическим обеспечением для разработки программного средства, ориентированного на медицинский персонал, позволяющего осуществлять поддержку принятия решений врача относительно пациентов с данным заболеванием.

Проведенный сравнительный анализ выявил, что разработанные на основе предложенного метода построения классификаторов модели обладают бóльшей точностью по сравнению с моделями, построенными на основе классических методов статистического моделирования (анализ дискриминантных функций, логистическая регрессия, метод деревьев решений). По показателям общей точности классификации превышение составляет $4,4 \div 8,9\%$, по специфичности – $4,2 \div 6\%$, по чувствительности – $7,6 \div 24,8\%$

Ключевые слова: классификация с обучением (с учителем), композиции классификаторов, модели оценки степени тяжести состояния пациента, информационная технология.

ABSTRACT

Nessonova M.N. Models and methods for patients' degree of severity assessment to support physician's decision. – Manuscript.

The thesis for candidate degree of technical science, specialty 05.13.09 – medical and biological informatics and cybernetics. – International Research and Training Center of Information Technology and Systems, National Academy of Sciences and Ministry of Science and Education of Ukraine, Kyiv, 2015.

The new decision of topical theoretical and applied task of models and methods development for patients' degree of severity assessment to support physician's decision is suggested in the thesis.

The new method for supervised classification is suggested. The method is based on the spatial representation analysis of relationships between classes and describing patterns, which is derived as a result of utilization of methods for geometrical interpretation of multidimensional data structure. Methods for classifiers compositions assembling have been further developed due to the combination of principles of specialization and weighted majority vote. The developed methods to design classifiers by learning information and their compositions were realized in the information technology, and this gives an ability to develop models for patients' degree of severity assessment and for clinical outcome prediction in cases of pancreas trauma and traumatic pancreatitis, bile-excreting ducts diseases, and ischemic and hemorrhagic strokes. The conducted comparative analysis shows that the developed models have higher accuracy in comparison with models, which were developed using classical methods of statistical modeling and analysis.

Keywords: supervised classification, classifiers compositions, models for patient's degree of severity estimation, information technology.