



розрізнити їх. Мікроскопія в лабораторних дослідженнях застосовується, або зі специфічними барвниками, або взагалі з флуоресціюючими антитілами, які прикріплюються до бактерій і таким чином роблять їх видимими.

У крові можна виявити яйця і личинки деяких гельмінтів. Наприклад, у шистосом є період гематогенної дисемінації, тобто поширення по організму з потоком крові. Ось тільки в периферичній крові їх виявити неможливо: занадто маленький діаметр у капілярів пальця, звідки беруть матеріал для аналізу. Яйця мають розміри 140-240 на 50-85 мкм. Середній розмір еритроцита - 7,5 мкм. Отже, це теж викликає певні сумніви.

За допомогою мікроскопа в краплі крові, взятої з пальця, не можна також побачити рН крові; дефіцит ферментів для розщеплення білків; рівень водно-сольового обміну; харчові мутагенні та тератогенні токсини; ураження еритроцитів нирковими токсинами та вільними радикалами; паразитів, гриби, бактерії, яйця глистів, цисти; активність, кількість і якість імунних клітин.

Безумовно дана методика має діагностичну цінність, поряд з традиційною мікроскопією мазка. Можна побачити, наприклад, серповидноклітинну анемію, та інші серйозні захворювання. Для наукових та клінічних досліджень надзвичайний інтерес представляє можливість дослідження та візуалізація процесів, які відбуваються в живій крові у динаміці, а саме під час руху її в капілярах.

**Іванчук М.А.**

### **МЕТОДИ ЛІНІЙНОЇ КЛАСИФІКАЦІЇ**

*Кафедра біологічної фізики та медичної інформатики*

*Вищий державний навчальний заклад України*

*«Буковинський державний медичний університет»*

Метою даної роботи є порівняння основних методів класифікації. Класифікація – один з розділів машинного навчання, що присвячений розв'язанню наступної задачі: існує множина об'єктів, розподілених певним чином на класи, кожен об'єкт характеризується фіксованою кількістю ознак. Задана навчальна вибірка, для об'єктів якої відомо, до яких класів вони відносяться. Класова приналежність інших об'єктів невідома. Необхідно побудувати алгоритм, здатний класифікувати довільний об'єкт з вихідної множини. Розглянемо основні методи лінійної класифікації.

**Лінійний дискримінантний аналіз** був запропонований Р.Фішером ще у 1936 році. Основна ідея методу полягає в тому, щоб вважати коваріаційні матриці класів рівними, навіть якщо вони насправді не є рівними. У такому випадку досить оцінити тільки одну коваріаційну матрицю. При цьому відокремлююча поверхня є лінійною, якщо класів два, і кусково-лінійною, якщо класів більше. Такий підхід збільшує стійкість класифікації, особливо у випадку малочисельних класів. Тому лінійний дискримінант часто виявляється кращим, ніж квадратичний, навіть в тих випадках, коли гіпотеза рівних коваріаційних матриць не вірна. Складність алгоритму лінійного дискримінантного аналізу  $O(nd+t^2)$ , де  $n$  – кількість спостережень в навчальній вибірці,  $d$  – кількість ознак,  $t = \min(n, d)$ .

**«Наївний» байєсівський класифікатор** спирається на додаткове припущення про статистичну незалежність ознак, що рідко виконується на практиці. Однак, завдяки простоті реалізації та ефективності, він найчастіше використовується як еталон для порівняння алгоритмів або як елементарний блок в більш складних моделях. Складність алгоритму наївного Байєсівського класифікатора складає  $O(nd)$ .

**Метод опорних векторів** завдяки принципу оптимальної відокремлюючої гіперплощини призводить до максимізації ширини відокремлюючої смуги між класами. Отже, цей метод призводить до якіснішої класифікації. Суть методу полягає у виділенні з усіх точок таких, що лежать на відстані  $1/||\omega||$  від відокремлюючої гіперплощини. Такі точки називаються опорними векторами. Суттєвою відмінністю методу від інших способів класифікації є те, що алгоритм методу не зміниться, якщо видалити всі точки, крім опорних. Суттєвим недоліком методу є відсутність розроблених загальних методів побудови спрямляючих просторів та ядер, що найкраще підходять до конкретної задачі. Складність алгоритму методу опорних векторів складає  $O(n^2)$ .

Для аналізу якості класифікації описані вище методи були нами застосовані для класифікації в медичних дослідженнях та на класичних базах даних. Наведемо приклад класифікації бази даних іриси Фішера, яка містить дані про 150 рослин трьох видів ірису (*setosa*, *versicolor*, *virginica*). Для кожної рослини відомо 4 ознаки – довжина і ширина чашолистика та довжина і ширина пелюстки. Необхідно за цими даними побудувати алгоритм класифікації. Ця база даних є цікавою тим, що перший клас (*setosa*) лінійно відокремлюється від двох інших (1-й випадок взаємного розташування множин), а класи *versicolor* та *virginica* між собою лінійно не відокремлюються (2-й випадок). Клас *setosa* був вірно розпізнаний всіма методами, при розпізнаванні класу *versicolor* найменшу кількість помилок (2 квітки з 50) одержали при використанні дискримінантного аналізу, при розпізнаванні класу *virginica* не отримали жодної помилки при використанні методу опорних векторів.

Отже, можна зробити висновок, що кожен з описаних методів лінійної класифікації має свої переваги та недоліки. Методи з більшою обчислювальною складністю (лінійний дискримінантний аналіз та метод опорних векторів) зазвичай дають кращий результат, проте вимагають більше комп'ютерного часу для його отримання. При цьому наївний байєсівський класифікатор легший у реалізації, проте на практиці зазвичай при використанні цього методу отримується більша кількість помилок.