

## **ЗАСТОСУВАННЯ МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ В ЗАДАЧІ ДІАГНОСТИКИ МІТОХОНДРІАЛЬНИХ ЗАХВОРЮВАНЬ**

**Борисов<sup>1</sup> А.П., Васильєва<sup>2</sup> О.В., Гречаніна<sup>2</sup> Ю.Б., Поворознюк<sup>1</sup> А.І., Філатова<sup>1</sup> Г.Є.**

*<sup>1</sup>Національний технічний університет «ХПІ», м. Харків*

*<sup>2</sup>Харківський спеціалізований медико-генетичний центр, м. Харків*

Задача діагностики включає етап побудови діагностичної моделі на основі аналізу вхідних даних, який складається в визначенні інформаційно значущих ознак захворювань. Вхідні дані надані Харківським спеціалізованим медико-генетичним центром для аналізу ознак, які необхідні для діагностики мітохондріальних захворювань. Особливості цих даних в тому, що вони вузькоспеціалізовані та можуть бути неповними. Серед них присутні групи показників біохімії крові та УЗД. Необхідно визначити, як ці групи взаємодіють одна з іншою.

Для вирішення поставленої задачі необхідно знайти та застосувати метод, який покаже зв'язок між групами показників. Основними вимогами до шуканого методу є знаходження зв'язку однієї множини змінних від другої множини та можливість використання методу для неповних даних.

Існує велика кількість методів для знаходження зв'язків між змінними. Серед них відомі методи кореляційного, регресійного аналізів, методи групового урахування аргументів, які не дозволяють знаходити зв'язок між множинами змінних, крім того вони не можуть бути застосовані до неповних вхідних даних. В якості визначення зв'язку між множинами змінних запропоновано використовувати розташування об'єктів в багатомірному просторі ознак. Якщо припустити, що групи ознак зв'язані між собою, то в багатомірному просторі ознак об'єкти будуть згруповані. Тому запропоновано використовувати методи кластерного аналізу для визначення груп об'єктів, які дозволяють зробити висновки щодо наявності зв'язків між множинами ознак. Серед всіх методів кластерного аналізу запропоновано використовувати метод «Сіоре», тому що він дозволяє виконувати кластеризацію об'єктів, які описуються різними підмножинами вхідного простору ознак. Основна ідея методу складається в максимізації глобального критерію розбиття об'єктів на кластери. При цьому коефіцієнт відштовхування регулює рівень подібності об'єктів всередині кластера та фінальну кількість кластерів.

В ході застосування методу визначено його ефективність для наданих вхідних даних, проаналізовано показники мітохондріальних захворювань та зроблено висновки щодо їх взаємодії.