

*Сімон К.А., студентка 3  
курсу спеціальності 122  
«Комп'ютерні науки»  
Нескородєва Т. В.,  
завідувач кафедри  
інформаційних технологій*

## **ВИКОРИСТАННЯ МЕТОДІВ СТАТИСТИЧНОГО НАВЧАННЯ ДЛЯ ВИРІШЕННЯ ПРОБЛЕМ У СФЕРІ ОХОРОНИ ЗДОРОВ'Я**

*Донецький національний університет імені В. Стуса, м. Вінниця*

Статистичне навчання - це галузь штучного інтелекту, що використовує алгоритми для знаходження закономірностей в даних. Одна з областей, в яких статистичне навчання може знайти широке застосування, - це сфера охорони здоров'я.

За допомогою методів статистичного навчання можна покращити діагностику, прогнозування розвитку захворювань, визначення ефективності лікування, виявлення аномалій та інші аспекти охорони здоров'я [1].

Давайте розглянемо кілька прикладів застосування статистичного навчання в цій галузі.

Застосування алгоритмів машинного навчання дозволяє досягти високої точності в прогнозуванні ризику виникнення захворювань. Наприклад, застосування алгоритму Random Forest у поєднанні з даними медичних досліджень може допомогти прогнозувати ризик виникнення раку грудей з точністю більше 90%.

Методи глибинного навчання, такі як нейронні мережі, знаходять широке застосування в аналізі медичних даних. Наприклад, нейронна мережа може бути навчена для автоматичного виявлення пухлин на зображеннях мозку.

Рекомендаційні системи можуть бути використані для покращення ефективності процесу лікування. Наприклад, рекомендаційна система може аналізувати медичні записи та попереджати лікарів про можливі побічні ефекти взаємодії між різними лікарськими препаратами.

Застосування алгоритмів кластеризації може допомогти виявити групи людей зі схожими медичними характеристиками та допомогти у плануванні профілактичних заходів. Наприклад, кластеризація може допомогти виявити групи людей з підвищеним ризиком розвитку цукрового діабету та зробити індивідуальний план профілактики для кожної групи [2].

Використання методів кластеризації та класифікації для прогнозування ризику виникнення захворювань. Наприклад, у дослідженні "Application of

machine learning algorithms for breast cancer prediction and diagnosis using clinical dataset" було використано алгоритми кластеризації та класифікації для прогнозування ризику виникнення раку молочної залози.

Використання методів обробки природних мов для аналізу медичних даних. Наприклад, у дослідженні "Application of natural language processing in health care: A literature review" було досліджено використання методів обробки природних мов для аналізу медичних даних, таких як медичні записи, листування з пацієнтами тощо [3].

Використання рекомендаційних систем для покращення якості надання медичної допомоги. Наприклад, у дослідженні "A novel collaborative filtering framework to personalize health recommendations" було розроблено рекомендаційну систему для персоналізації рекомендацій щодо здорового способу життя.

Використання методів аналізу соціальних мереж для вивчення поведінки пацієнтів та покращення надання медичної допомоги. Наприклад, у дослідженні "Social media and health care: A systematic review" було досліджено використання соціальних мереж для аналізу поведінки пацієнтів, зокрема, щодо застосування лікарських засобів.

Використання методів обробки зображень для діагностики захворювань. Наприклад, у дослідженні "Deep learning for medical image analysis: A review" було досліджено використання глибокого навчання для обробки зображень, зокрема, для діагностики захворювань, таких як рак молочної залози [4].

Також, існують методи статистичного навчання, які можуть допомогти у виявленні різноманітних хвороб та їх прогнозуванні. Наприклад, методи класифікації, такі як логістична регресія та дерева рішень, можуть бути використані для класифікації пацієнтів за наявністю хвороби. Крім того, методи кластеризації можуть допомогти у виявленні груп пацієнтів зі схожими характеристиками, що може допомогти у визначенні ефективних методів лікування.

Деякі програмні реалізації методів статистичного навчання, які можуть бути використані в галузі охорони здоров'я, включають в себе R, Python, SAS та MATLAB. Наприклад, R та Python є популярними мовами програмування для статистичного аналізу та машинного навчання, тому їх можна використовувати для розробки різноманітних алгоритмів для аналізу даних в галузі охорони здоров'я. SAS та MATLAB також мають велику кількість функцій для аналізу даних та моделювання.

Нарешті, існують готові рекомендаційні системи, які можуть допомогти лікарям та пацієнтам у виборі ефективних методів лікування. Однією з таких систем є "IBM Watson for Oncology", яка використовує методи машинного навчання для аналізу медичних даних та рекомендацій лікування пацієнтам з онкологічними захворюваннями. Ця система використовує базу даних, що містить понад 300 тисяч статей та понад 200 тисяч клінічних записів, для аналізу симптомів та лікування [5].

Мова R є популярним інструментом в галузі охорони здоров'я, оскільки вона надає засоби для проведення статистичного аналізу даних та побудови моделей машинного навчання. Деякі з основних областей застосування мови R в охороні здоров'я включають:

Епідеміологія та дослідження клінічних випадків: Мова R використовується для статистичного аналізу даних, отриманих з клінічних випадків, досліджень та епідеміологічних досліджень. Наприклад, в R можна проводити аналізи виживання, побудову кривих ризику та різних моделей прогнозування.

Геноміка та біоінформатика: Мова R використовується для аналізу геномних даних та встановлення зв'язків між генами та хворобами. Вона також допомагає у розробці програмного забезпечення для обробки генетичних даних та побудови моделей класифікації.

Системи здоров'я та страхування: Мова R використовується для аналізу даних з систем охорони здоров'я та страхування. Наприклад, можна використовувати R для аналізу даних щодо ефективності лікування та планування фінансових витрат на охорону здоров'я.

Інформаційні технології в охороні здоров'я: Мова R використовується для розробки програмного забезпечення, яке допомагає в рішенні проблем, пов'язаних зі збереженням та обробкою медичних даних. Наприклад, R використовується для розробки рекомендаційних систем, які надають лікарям рекомендації щодо лікування [6].

Застосування мови R в галузі охорони здоров'я може бути досить різноманітним. Нижче наведені декілька прикладів:

Аналіз клінічних даних: мова R широко використовується для аналізу клінічних даних. Наприклад, за допомогою пакетів `dplyr` та `ggplot2` можна здійснювати різноманітний аналіз даних, такий як обчислення статистичних показників (середні, медіани, стандартні відхилення), побудову графіків, в тому числі боксплотів, стовпчикових діаграм, графіків розкиду тощо. Крім того, за допомогою пакетів `survival` та `coxph` можна здійснювати аналіз даних з використанням методів виживання, що дозволяє вивчати час виникнення подій, таких як смерть або рецидив захворювання.

Моделювання епідемій: мова R може бути використана для моделювання епідемій із використанням методу стохастичного моделювання. Наприклад, пакетів `EpiModel` та `romp` дозволяють моделювати поширення інфекційних хвороб, таких як грип, COVID-19, ВІЛ/СНІД та інші. Ці пакети містять інструменти для побудови математичних моделей, оцінки параметрів моделей та аналізу даних.

Аналіз зображень: мова R може бути використана для аналізу медичних зображень. Наприклад, пакетів `EImage` та `radiomics` дозволяють зчитувати, обробляти та аналізувати медичні зображення, такі як зображення мозку, легенів, серця тощо. За допомогою цих пакетів можна здійснювати аналіз зображень, включаючи розпізнавання об'єктів [7].

### Список літератури:

1. Альгамді, М., і Хан, В. (2019). *Машинне навчання для охорони здоров'я: огляд, можливості та проблеми. У матеріалах 9-ї міжнародної конференції з хмарних обчислень і великих даних (стор. 82-91). АСМ.*
2. Вікхем, Х. (2016). *ggplot2: елегантна графіка для аналізу даних. Springer-Verlag Нью-Йорк.*
3. Кун, М. (2019). *caret: навчання класифікації та регресії. Пакет R версії 6.0-84.*
4. Леделл Е., Петерсен М. Л. та ван дер Лаан М. Дж. (2016). *Обчислювально ефективні довірчі інтервали для перехресно перевіреної області під оцінками кривої ROC. Електронний журнал статистики, 10(1), 1-22.*
5. Harrell Jr, F. E. (2015). *Стратегії регресійного моделювання: із застосуванням до лінійних моделей, логістичної регресії та аналізу виживання. Спрингер.*
6. Кун, М., Джонсон, К. (2013). *Прикладне прогнозне моделювання. Спрингер.*
7. Ю.С. Хмелівський, Т.В. Нескородева *Аналіз даних для прогнозування серцевої недостатності засобами мови R. Матеріали II всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних». Вінниця, Донецький національний університет імені Василя Стуса. С. 57-60. URL: <https://jktod.donnu.edu.ua/article/view/11631>*
8. В.М. Чернега Т.В. Нескородева *Аналіз факторів, що призводять до серцево-судинних захворювань. Матеріали II всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних». Вінниця, Донецький національний університет імені Василя Стуса. С. 60-63. URL: <https://jktod.donnu.edu.ua/article/view/11632>*