

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Навчально-науковий центр заочної форми навчання

(повна назва)

Кафедра

Інформаційних управляючих систем

(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА

рівень вищої освіти другий (магістерський)

Дослідження методів підтримки діагностичних рішень  
у медичних інформаційних системах на основі мутаційного аналізу

(тема)

Виконав:

здобувач 2 року навчання,

групи ІУСТзм-23-1

Чалкова Дар'я Дмитрівна

(прізвище, ім'я, по батькові)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні управляючі системи та технології

(повна назва освітньої програми)

Керівник: проф. каф. ІУС Чала О. В.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ІУС



(підпис)

Петров К.Е.

(прізвище, ініціали)

2025 р.

## Харківський національний університет радіоелектроніки

Навчально-науковий центр заочної форми навчання  
 Кафедра Інформаційних управляючих систем  
 Рівень вищої освіти другий (магістерський)  
 Спеціальність 122 Комп'ютерні науки  
 (код і повна назва)  
 Тип програми освітньо-професійна  
 (освітньо-професійна або освітньо-наукова)  
 Освітня програма Інформаційні управляючі системи та технології  
 (повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри

(підпис)

« 9 » грудня 20 24 р.

## ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Чалковій Дар'ї Дмитрівні  
 (прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу

затверджена наказом по університету від « 03 » грудня 2024 р. № 205Стз

2. Термін подання студентом роботи до екзаменаційної комісії « 14 » січня 2025 р.

3. Вихідні дані до роботи Науково-технічні публікації та інтернет джерела з тематики кваліфікаційної роботи

4. Перелік питань, що потрібно опрацювати в роботі аналіз предметної області та постановка задачі дослідження, дослідження методів підтримки діагностичних рішень у медичних інформаційних системах, розробка підходу підтримки діагностичних рішень на основі мутаційного аналізу, впровадження модуля прогнозування ризиків та планування терапії з використанням CRISPR, інтеграція фреймворку в електронну медичну карту, експериментальна перевірка роботи удосконаленого методу та його ефективності, висновки.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів	Примітка
1	Аналіз предметної області	09.12.2024 – 15.12.2024	Виконано
2	Постановка задачі	16.12.2024-17.12.2024	Виконано
3	Обробка матеріалу	17.12.2024-22.12.2024	Виконано
4	Дослідження методів підтримки рішень	23.12.2024-24.12.2024	Виконано
5	Дослідження медичних інформаційних систем	25.12.2024-27.12.2024	Виконано
6	Експериментальна перевірка	28.12.2024-29.12.2024	Виконано
7	Аналіз результатів дослідження	30.12.2024-03.01.2025	Виконано
8	Написання пояснювальної записки	04.01.2025-11.01.2025	Виконано
9	Підготовка презентації	06.01.2025-14.01.2025	Виконано
10	Перевірка на плагіат	15.01.2025	Виконано
11	Нормконтроль	16.01.2025	Виконано
12	Захист	17.01.2025	Виконано

Дата видачі завдання 09 грудня 2024 р.

Студент \_\_\_\_\_

(підпис)

Керівник роботи \_\_\_\_\_

(підпис)

проф. каф. ІУС Чала О.В.

(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 113 с., 18 рис., 3 табл., 1 дод., 46 джерел.

**ІНТЕГРАЦІЯ В ЕЛЕКТРОННІ МЕДИЧНІ КАРТИ, МУТАЦІЙНИЙ АНАЛІЗ, ПІДТРИМКА ДІАГНОСТИЧНИХ РІШЕНЬ, ПЛАНУВАННЯ ТЕРАПІЇ, ПРОГНОЗУВАННЯ РИЗИКІВ.**

У роботі проведено дослідження методів підтримки діагностичних рішень у медичних інформаційних системах на основі аналізу мутацій. Визначено ключові вимоги до інтеграції геномних даних у процеси прийняття рішень у персоналізованій медицині, розроблено концепцію інформаційної технології, орієнтованої на автоматизацію та прискорення клінічної діагностики. Проведено аналіз існуючих підходів до мутаційного аналізу та інструментів для його реалізації, таких як SComatic.

Розроблений фреймворк включає кілька основних модулів: діагностичний інструмент для оцінки мутацій, систему прогностичної оцінки ризиків, планувальник терапії на основі CRISPR та механізми інтеграції з електронними медичними картами. Методологія ґрунтується на адаптації сучасних алгоритмів аналізу мутацій і їхньої інтерпретації для виявлення діагностично значущих змін у геномі.

Особливу увагу приділено технологічним аспектам реалізації, включаючи вибір алгоритмів машинного навчання для прогнозування, структурування баз даних для геномних даних, а також забезпечення сумісності фреймворку з існуючими медичними інформаційними системами. Розроблена система дозволяє підвищити точність діагностики, зменшити кількість помилкових висновків та створити основу для персоналізованої терапії.

## ABSTRACT

Master's thesis: 113 pages, 18 figures, 3 tables, 1 appendix, 46 sources.

DIAGNOSTIC DECISION SUPPORT, INTEGRATION INTO ELECTRONIC MEDICAL RECORDS, MUTATIONAL ANALYSIS, RISK PREDICTION, THERAPY PLANNING.

This study explores methods for supporting diagnostic decision-making in medical information systems based on mutational analysis. Key requirements for integrating genomic data into decision-making processes in personalized medicine were identified, and a concept for an information technology framework focused on automation and acceleration of clinical diagnostics was developed. An analysis of existing approaches and tools for mutational analysis, such as SComatic, was conducted.

The developed framework includes several core modules: a diagnostic tool for mutation evaluation, a risk prediction system, a therapy planning tool based on CRISPR, and mechanisms for integration with electronic medical records. The methodology is based on adapting modern algorithms for mutation analysis and interpretation to identify diagnostically significant genomic changes.

Particular attention was paid to the technological aspects of implementation, including the selection of machine learning algorithms for risk prediction, structuring genomic databases, and ensuring the framework's compatibility with existing medical information systems. The proposed system improves diagnostic accuracy, reduces the incidence of false conclusions, and establishes a foundation for personalized therapy.

## ЗМІСТ

	С.
Скорочення та умовні позначки .....	8
Вступ.....	10
1 Аналіз предметної області та постановка задачі дослідження.....	12
1.1 Аналіз характеристик медичних інформаційних систем.....	12
1.2 Дослідження процесів діагностування з використанням мутаційного аналізу.....	19
1.3 Аналіз методів підтримки діагностичних рішень у медичних інформаційних системах .....	29
1.4 Постановка задачі дослідження.....	34
2 Підтримка діагностичних рішень у медичних інформаційних системах з використанням мутаційного аналізу .....	38
2.1 Розробка підходу до підтримки діагностичних рішень на основі мутаційного аналізу .....	38
2.2 Розробка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу.....	52
3 Розробка інформаційної технології підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу .....	54
3.1 Концепція та вимоги до розробки технології.....	54
3.2 Архітектура фреймворку на основі SComatic .....	57
3.3 Технологічні аспекти реалізації.....	64
4 Експериментальна перевірка отриманих наукових результатів .....	71
4.1 Розробка модулю підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу.....	71
4.2 Експериментальна перевірка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу.....	86

Висновки .....	92
Перелік джерел посилання .....	94
Додаток А Графічний матеріал кваліфікаційної роботи.....	101

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ДНК – дезоксирибонуклеїнова кислота

ІТ – інформаційні технології

МІС – медичні інформаційні системи

РНК – рибонуклеїнова кислота

СППР – системи підтримки прийняття рішень

ПЗ – програмне забезпечення

ТЗ – технічне завдання

APC – ген аденоматозного поліпозу товстої кишки

ATM – ген, який кодує білок, що бере участь у виправленні пошкоджень ДНК

BRAF – ген, що кодує білок кіназу в сигнальних шляхах

COSMIC – Catalogue Of Somatic Mutations In Cancer

CNV – варіація числа копій (Copy Number Variation)

CRISPR – напрямлене редагування геному за допомогою системи CRISPR/Cas9

FASTA – формат асоційованих строк

FASTAq – розширення формату FASTA для додавання інформації про якість

GISTIC – генетичний інструмент для ідентифікації ампліфікацій і делецій в геномі

gRNA – напрямлена (гайдова) РНК для системи CRISPR

NGS – технологія секвенування нового покоління

p53 – ген, що відповідає за регуляцію клітинного циклу та апоптоз

SNP – однонуклеотидний поліморфізм

TP53 – третій алель гена p53, що несе важливі функції в онкології

VAF – частка алелей візуалізованих з мутацією (Variant Allele Frequency)

WES – секвенування екзому

WGS – секвенування всього геному

## ВСТУП

У сучасному світі медичної інформатики ключовою стає розробка технологій підтримки прийняття діагностичних рішень, що забезпечують ефективне управління зростаючим обсягом медичних даних, зокрема геномної інформації. Такі технології спрямовані на персоналізацію діагностичних підходів та лікування, забезпечуючи високу точність і оперативність у процесі прийняття клінічних рішень.

Ефективне впровадження мутаційного аналізу в медичну практику потребує створення інструментів, здатних не лише інтегрувати геномні дані з іншими джерелами медичної інформації, а й підтримувати прийняття обґрунтованих рішень у діагностиці та терапії. Це вимагає застосування сучасних методів обробки даних і забезпечення їхньої сумісності з існуючими медичними інформаційними системами.

Об'єктом дослідження є медичні інформаційні системи, які використовують мутаційний аналіз для підтримки діагностичних рішень.

Предметом дослідження виступають методи підтримки прийняття діагностичних рішень на основі аналізу мутацій у медичних інформаційних системах.

Метою дослідження є удосконалення методу для підтримки діагностичних рішень, яке дозволить автоматизувати аналіз мутаційних даних, оцінювати ризики та планувати персоналізовану терапію.

Задачі дослідження:

- провести аналіз предметної області;
- розробити підхід до підтримки діагностичних рішень на основі мутаційного аналізу;
- удосконалити метод формування набору діагностичних рішень з урахуванням обмежень за результатами мутаційного аналізу;

-розробити інформаційну технологію підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу;

-провести експериментальну перевірку отриманих наукових результатів.

Наукова новизна дослідження полягає в удосконаленні метода підтримки діагностичних рішень, який поєднує мутаційний аналіз із прогностичними та терапевтичними функціями, адаптованими до персоналізованої медицини.

Практична значущість роботи полягає в можливості застосування розробленої системи у клінічній практиці для підвищення ефективності та якості медичних послуг за рахунок ефективною підтримки прийняття діагностичних рішень.

Робота пройшла апробацію дослідження:

1) Чалкова Д. Д. Дослідження методів підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу.

III Міжнародна наукова конференція «Технології та суспільство: взаємодія, вплив, трансформація», 17 січня 2025 р., Львів, Україна. = Proceedings of the III International Scientific Conference «Technologies and Society: Interaction, Influence, Transformation» (January 17, 2025), Lviv, Ukraine. Рр. 456–461. Доступно за посиланням: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/17.01.2025>;

2) Чалкова Д. Д. Дослідження методів підтримки діагностичних рішень у медичних інформаційних системах. VII Міжнародна студентська наукова конференція «Модернізація та сучасні українські і світові наукові дослідження», 24 січня 2025 р., Луцьк, Україна. = Proceedings of the VII International Student Scientific Conference «Modernization and Current Ukrainian and World Scientific Research» (January 24, 2025), Lutsk, Ukraine. Рр. 320–326. Доступно за посиланням: <https://archive.liga.science/index.php/conference-proceedings/issue/view/inter-24.01.2025>.

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

## 1.1 Аналіз характеристик медичних інформаційних систем

Медичні інформаційні системи (МІС) відіграють ключову роль у сучасній медицині, забезпечуючи комплексну обробку даних пацієнтів, управління процесами діагностики, лікування та досліджень. Їхній розвиток зумовлений необхідністю інтеграції великого обсягу різномірних даних із різних джерел, таких як клінічні записи, лабораторні дослідження, результати візуалізації, геномні дані, а також дані з електронних пристроїв пацієнтів. У сучасних умовах важливість таких систем зростає через зростання обсягу даних у медицині та необхідність їх швидкої обробки.

Ключова мета МІС полягає у створенні цілісного підходу до управління інформацією, що забезпечує доступ лікарів до всіх релевантних даних у реальному часі. Наприклад, системи електронних медичних записів дозволяють зберігати не лише текстову інформацію про пацієнта, а й зображення, такі як рентгенівські знімки або гістологічні зрізи. Це спрощує доступ до результатів досліджень і дозволяє лікарям швидше ухвалювати обґрунтовані рішення.

Окрім основного завдання зі збереження даних, МІС інтегрують інструменти для аналізу інформації. Зокрема, системи підтримки прийняття рішень (CDSS) використовуються для діагностики захворювань і розробки індивідуальних планів лікування [1]. Наприклад, сучасні системи можуть ідентифікувати потенційні взаємодії між лікарськими засобами або оцінювати ризик розвитку ускладнень, використовуючи алгоритми на основі великих даних [2]. У контексті аналізу мутацій, такі системи допомагають інтерпретувати дані секвенування генома пацієнтів, визначаючи мутації, що можуть бути причиною захворювань, і пропонуючи рекомендації для подальших клінічних дій.

Технологічна складова МІС включає використання складних баз даних, які

адаптовані для зберігання великого обсягу даних різних типів. Наприклад, геномні дані часто зберігаються у спеціалізованих форматах, сумісних із технологіями великих даних. Це дозволяє інтегрувати різні джерела інформації, забезпечуючи комплексний аналіз. Особлива увага приділяється питанням інтеперабельності, тобто можливості обміну даними між різними системами. Для цього використовуються міжнародні стандарти, такі як HL7 і FHIR, що забезпечують ефективний обмін даними між лікарнями, лабораторіями, аптеками та дослідницькими центрами [3].

Одним із ключових аспектів МІС є забезпечення безпеки даних, особливо враховуючи їх конфіденційний характер. У сучасних системах впроваджуються технології шифрування та багаторівнева автентифікація, які мінімізують ризики несанкціонованого доступу. Це дозволяє дотримуватися міжнародних стандартів захисту даних, таких як GDPR та HIPAA, що є обов'язковими для медичних установ у багатьох країнах [4].

Класифікація медичних інформаційних систем (МІС) охоплює різноманітні аспекти їхнього використання, функціональності та впровадження. Вона дозволяє структурувати знання про ці системи, враховуючи їх призначення, масштаб, інтеграцію та рівень впливу на медичну практику. Схема нижче (рис. 1.1) ілюструє основні критерії класифікації МІС, підсистеми діагностування та їх взаємозв'язок із іншими складовими системами.

За функціональним призначенням МІС поділяються на кілька основних груп. До систем управління належать ті, що орієнтовані на адміністрування медичних установ, включаючи реєстрацію пацієнтів, облік персоналу, управління фінансами та забезпечення ресурсами. Клінічні системи призначені для підтримки медичного обслуговування, таких як введення і зберігання електронних медичних записів, управління процесами діагностування, лікування та реабілітації пацієнтів. Науково-дослідні системи використовуються для зберігання, аналізу та моделювання даних з метою підтримки дослідницької діяльності, зокрема у фармакології, геноміці та епідеміології. Інформаційні

системи телемедицини забезпечують віддалене консультування пацієнтів, моніторинг стану здоров'я та діагностику в режимі реального часу, що є особливо важливим для віддалених регіонів.

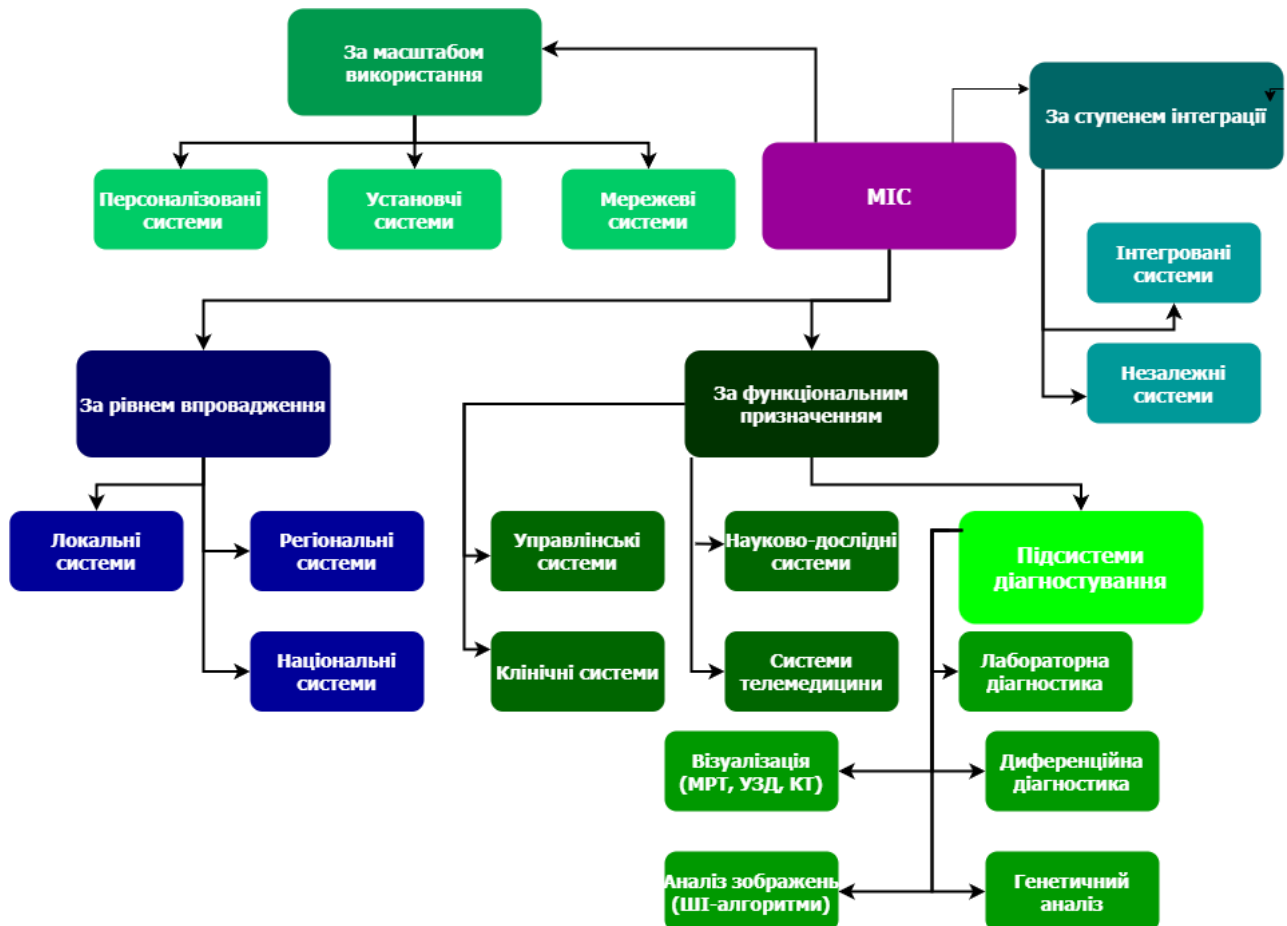


Рисунок 1. 1 – Класифікація МІС

Підсистеми діагностування є ключовим елементом МІС, оскільки вони забезпечують автоматизацію і точність у постановці діагнозів. Ці підсистеми можуть включати спеціалізовані модулі, що аналізують дані, отримані від різних джерел, таких як лабораторні тести, рентгенівські знімки, результати УЗД або МРТ. Використання алгоритмів штучного інтелекту дозволяє значно підвищити швидкість і точність діагностики, наприклад, за рахунок автоматичного виявлення патологій на зображеннях. Інтеграція з базами даних про попередні захворювання, генетичні ризики та сучасні протоколи лікування сприяє

персоналізованому підходу до кожного пацієнта. Такі системи можуть також підтримувати функції диференційної діагностики, дозволяючи лікарям аналізувати симптоми, оцінювати ймовірність певних захворювань та вибирати найбільш ефективну стратегію лікування.

За рівнем впровадження МІС класифікують на локальні, регіональні та національні системи. Локальні системи використовуються окремими медичними установами і забезпечують підтримку внутрішніх процесів, таких як реєстрація пацієнтів, управління графіками лікарів, облік витрат матеріалів і медикаментів. Регіональні системи охоплюють кілька закладів охорони здоров'я в межах одного регіону і часто інтегруються з іншими службами, такими як лабораторії, аптеки або швидка допомога. Національні системи працюють на рівні всієї країни, об'єднуючи дані з усіх медичних установ та забезпечуючи централізований моніторинг епідеміологічної ситуації, планування медичних ресурсів і впровадження державних програм.

За ступенем інтеграції медичні інформаційні системи поділяються на незалежні та інтегровані. Незалежні системи функціонують автономно і виконують певні задачі, наприклад, лабораторні інформаційні системи або модулі управління запасами. Інтегровані системи об'єднують різні компоненти в єдину платформу, забезпечуючи взаємодію між підсистемами діагностики, лікування, адміністрування та управління ресурсами. Наприклад, інтегрована система може автоматично передавати результати лабораторних досліджень у електронну медичну карту пацієнта або синхронізувати дані з модулем призначень лікаря [3, 5].

Таким чином, медичні інформаційні системи є багатограними інструментами, які значно підвищують ефективність медичних послуг, оптимізують процеси обробки даних та сприяють покращенню якості лікування. Їхнє впровадження потребує зваженого підходу, враховуючи специфіку медичної практики, технічні можливості та потреби конкретних закладів охорони здоров'я. Особливу увагу слід приділяти розвитку підсистем

діагностування, які є основою для швидкого, точного та ефективного надання медичних послуг.

Незважаючи на очевидні переваги, впровадження МІС стикається з низкою викликів. Одним із найбільш актуальних є стандартизація даних, оскільки різні установи можуть використовувати несумісні формати. Крім того, масштабування систем стає важким завданням через постійне зростання обсягу медичних даних. Це вимагає впровадження нових підходів до оптимізації алгоритмів обробки, а також удосконалення апаратного забезпечення [6].

Архітектура МІС, представлена на рисунку 1.2, базується на багаторівневій моделі, яка забезпечує ефективне управління даними, логікою бізнес-процесів та взаємодією з користувачем, а також інтеграцію механізмів безпеки [3, 12].

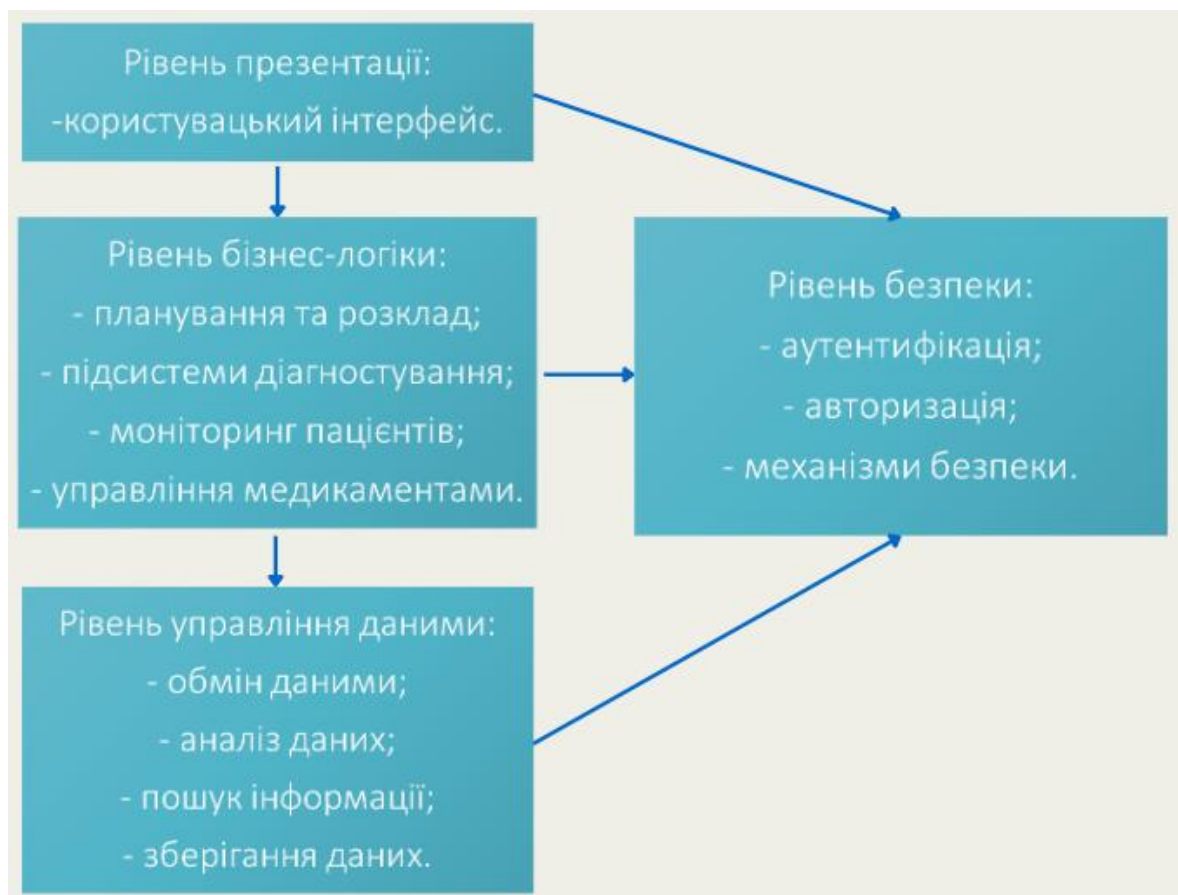


Рисунок 1.2 – Архітектура МІС

Верхній рівень — це рівень презентації, який відповідає за взаємодію

користувача з системою через інтерфейс. Цей рівень забезпечує зручний доступ до функціоналу та передачу даних до інших рівнів.

Рівень бізнес-логіки забезпечує виконання основних функцій МІС, таких як планування, моніторинг пацієнтів, управління медикаментами та інтеграція загальних функцій медичних інформаційних систем. Цей рівень координує дії між користувачем та рівнем управління даними, а також здійснює доступ до механізмів безпеки, таких як автентифікація та авторизація.

Нижній рівень — рівень управління даними — відповідає за обробку, аналіз та зберігання даних. Він включає функції аналітики, обміну даними, пошуку інформації, інтеграції даних із сенсорів та надання допомоги. Цей рівень є основою для обробки та зберігання великих обсягів медичних даних, забезпечуючи можливість швидкого доступу та аналізу.

Важливу роль у цій архітектурі відіграє шар безпеки, який працює на всіх рівнях системи. Він включає механізми автентифікації, авторизації та захисту даних. Цей шар гарантує, що доступ до системи отримують лише авторизовані користувачі, а передача та обробка даних залишаються захищеними від несанкціонованого доступу. Таким чином, представлена архітектура забезпечує модульність, функціональність та безпечність медичної інформаційної системи.

З одного боку, МІС можуть бути корисними для полегшення обробки та інтерпретації результатів мутаційного аналізу, зокрема завдяки впровадженню методів машинного навчання та штучного інтелекту, які здатні виявляти патерни в даних, які можуть бути не очевидними для людини [13]. Наприклад, ШІ може допомогти виявляти генетичні маркери, які пов'язані з підвищеним ризиком розвитку певних захворювань, таких як рак, що дозволяє створювати більш точні моделі для прогнозування та планування лікування [14].

Інший приклад використання ШІ в МІС, системи на основі глибокого навчання, такі як Zebra Medical Vision або Aidoc, аналізують медичні зображення (рентгенівські знімки, КТ чи МРТ), автоматично виявляючи патології, включаючи пухлини чи тромби, з точністю, що конкурує з результатами

досвідчених радіологів [15]. У галузі геноміки ШІ дозволяє проводити швидку інтерпретацію генетичних варіацій, визначаючи ризики спадкових захворювань. Так, система DeepVariant від Google застосовує глибокі нейронні мережі для покращення точності виявлення мутацій на основі секвенування ДНК [14].

Крім діагностики, ШІ в МІС може підтримувати управління терапевтичними планами. Наприклад, IBM Watson for Oncology надає лікарям рекомендації щодо лікування раку, ґрунтуючись на останніх наукових даних і клінічних протоколах. Такий підхід не лише зменшує час на вибір терапії, а й дозволяє адаптувати лікування до індивідуальних потреб пацієнта. Іншим прикладом є застосування ШІ для оптимізації логістики та управління медичними ресурсами. Алгоритми, інтегровані в МІС, можуть прогнозувати потребу в медикаментах чи обладнанні на основі історичних даних та поточного стану пацієнтів у лікарнях [16].

Інтеграція ШІ також сприяє підвищенню безпеки пацієнтів. Наприклад, автоматизовані системи, такі як MedAware, аналізують призначення лікарських засобів, попереджаючи про можливі помилки, несумісності або потенційні побічні ефекти. Крім того, використання алгоритмів обробки природної мови в МІС дозволяє швидше витягувати важливу інформацію з незструктурованих текстів, таких як клінічні звіти чи записи анамнезу, прискорюючи процес ухвалення рішень лікарями.

Додатково, на ефективність МІС може впливати й фактор людського чинника, зокрема рівень кваліфікації медичного персоналу, що працює з системою. У багатьох випадках лікарі стикаються з труднощами в інтерпретації складних геномних даних, навіть за наявності сучасних технологій, що вимагає додаткових навчальних програм і підвищення кваліфікації медичних працівників [17].

Отже, медичні інформаційні системи є невід'ємною частиною сучасної медицини, яка дозволяє автоматизувати процеси обробки даних, підвищити ефективність діагностики та лікування, а також забезпечити розвиток

персоналізованої медицини. Їхній розвиток базується на інтеграції сучасних технологій, таких як машинне навчання, хмарні обчислення та інтероперабельні стандарти, що відкриває нові можливості для клінічної практики та наукових досліджень.

## 1.2 Дослідження процесів діагностування з використанням мутаційного аналізу

Мутаційний аналіз є потужним інструментом, що знаходиться на перетині біології, медицини та комп'ютерних наук. Його метою є систематичне виявлення і інтерпретація змін у геномі, які впливають на біологічні функції або підвищують ризик розвитку захворювань. У сфері комп'ютерних наук цей процес включає розробку та впровадження складних алгоритмів, обчислювальних платформ та інструментів аналізу даних для ефективного виявлення, класифікації та прогнозування впливу мутацій.

Основна мета мутаційного аналізу полягає в автоматизації обробки великих обсягів геномних даних, отриманих із сучасних методів секвенування. Це дозволяє скоротити час на аналіз, підвищити точність і відтворюваність результатів, а також забезпечити інтеграцію цих даних із клінічними системами. З точки зору комп'ютерних наук, важливою метою є створення масштабованих систем, які можуть обробляти ексабайти генетичної інформації, враховуючи складність біологічних структур і можливість різних варіантів мутацій, таких як точкові зміни, інсерції, делеції або складні перебудови.

Необхідність мутаційного аналізу продиктована кількома ключовими факторами. Перш за все, кількість геномних даних стрімко зростає, що створює виклики для традиційних методів обробки. Алгоритми машинного навчання, глибокі нейронні мережі та статистичні моделі дозволяють значно підвищити

ефективність аналізу, виявляючи закономірності, які неможливо виявити вручну. Для реалізації цього потрібні обчислювальні платформи з високою продуктивністю, здатні працювати з розподіленими базами даних та оптимізованими алгоритмами [16].

По-друге, важливим аспектом є розробка інструментів для прогнозування впливу мутацій. Наприклад, зміни в геномі можуть мати як нейтральний, так і патогенний характер, а також змінювати взаємодії між генами та білками. Для моделювання цих взаємодій використовуються графові алгоритми, які можуть виявляти зв'язки між елементами генетичних мереж. Крім того, застосовуються моделі, які дозволяють оцінювати функціональні наслідки мутацій, такі як PolyPhen, SIFT або інші методи предиктивного аналізу, що потребують значної обчислювальної потужності [7].

Іншою важливою метою є інтеграція мутаційного аналізу з іншими джерелами даних, такими як клінічні записи, медичні зображення або екологічні фактори. Це створює додаткові виклики для комп'ютерних наук, пов'язані з інтеперабельністю систем, розробкою стандартів обміну даними (наприклад, формат VCF для геномних варіантів) та побудовою розподілених обчислювальних платформ [18].

Необхідність використання мутаційного аналізу також зумовлена потребою у персоналізованій медицині. Сучасні алгоритми дозволяють зіставляти генетичні профілі пацієнтів із наявними базами даних, визначати асоціації мутацій із певними захворюваннями та розробляти персоналізовані стратегії лікування (рис. 1.3).

Це вимагає побудови систем, які можуть адаптуватися до різних наборів даних та автоматично оновлювати свої знання на основі нової інформації.

Мутаційний аналіз відіграє ключову роль у сучасній медицині, зокрема в діагностуванні різноманітних генетичних захворювань та онкологічних патологій. Цей метод базується на вивченні змін у ДНК, що можуть впливати на функціонування клітин і спричиняти розвиток хвороб. Завдяки швидкому

прогресу технологій секвенування, зокрема методу нового покоління (NGS), стало можливим досліджувати геном людини з безпрецедентною точністю, що значно розширило можливості діагностування.

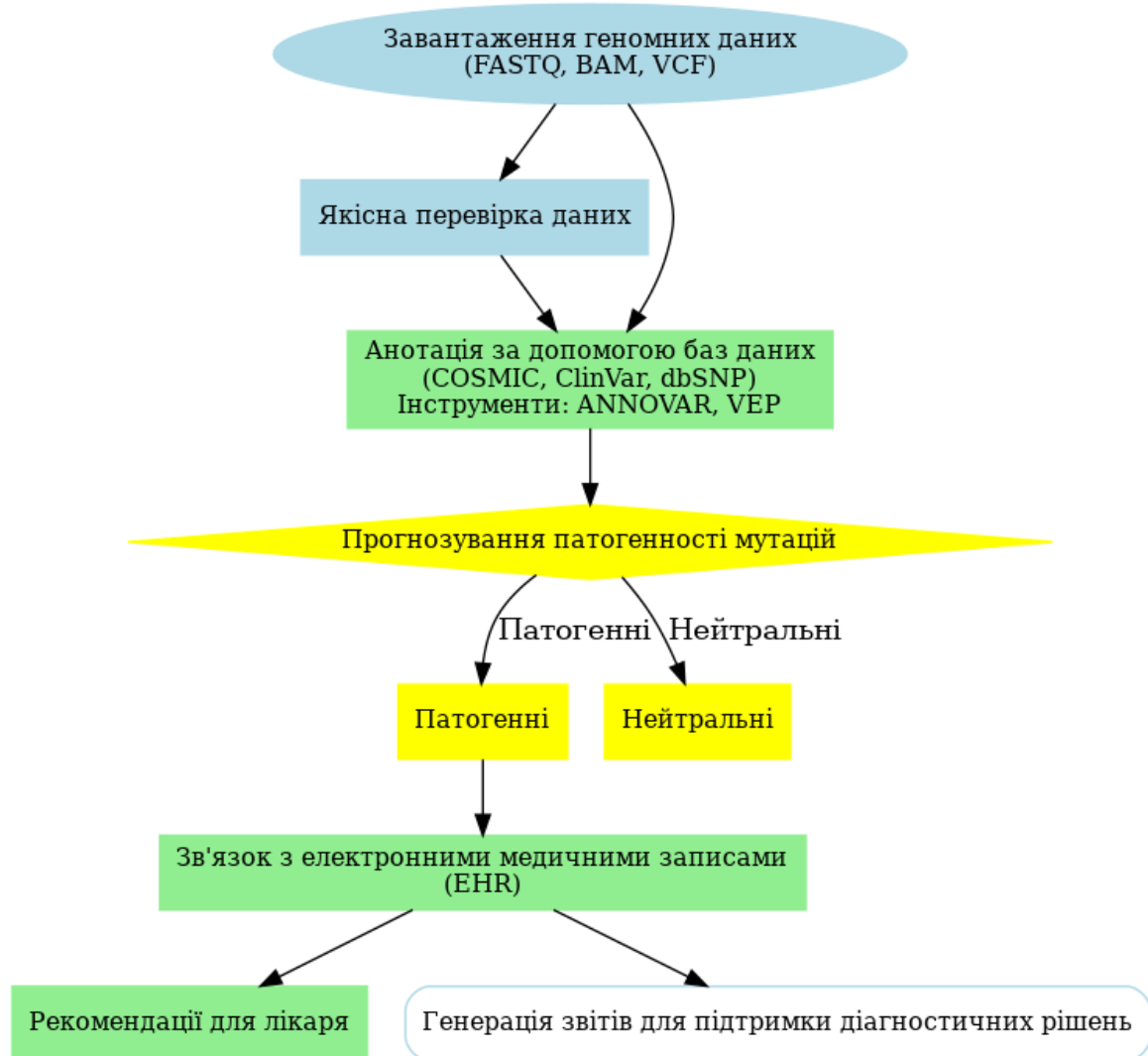


Рисунок 1.3 – Етапи процесу мутаційного аналізу в медичних інформаційних системах

Одним із найпоширеніших прикладів використання мутаційного аналізу є діагностика спадкових онкологічних синдромів, таких як синдром Лінча або BRCA-асоційований рак молочної залози та яєчників. У таких випадках аналіз спрямований на виявлення мутацій у певних генах, наприклад, *BRCA1* або *BRCA2*, які асоціюються з підвищеним ризиком розвитку раку. Виявлення таких

мутацій дозволяє не лише підтвердити діагноз, але й забезпечити персоналізований підхід до лікування та профілактики, включаючи регулярний моніторинг або профілактичну хірургію [19].

Ще одним важливим напрямком є використання мутаційного аналізу в онкогематології. Наприклад, дослідження мутацій у генах *FLT3* або *NPM1* дозволяє класифікувати пацієнтів із гострим мієлоїдним лейкозом на різні прогностичні групи, що впливає на вибір терапевтичної стратегії. Завдяки аналізу можна визначити, чи потребує пацієнт інтенсивнішого лікування, включаючи трансплантацію кісткового мозку [20].

Мутаційний аналіз також є невід'ємною частиною діагностики рідкісних генетичних захворювань, таких як муковісцидоз, синдром Марфана чи хвороба Гантінгтона. У таких випадках аналіз допомагає не лише встановити точний діагноз, але й оцінити ризик передачі хвороби у спадок, що є важливим аспектом для планування сім'ї [21].

Окрім діагностики, мутаційний аналіз знаходить широке застосування у фармакогенетиці. Вивчення генетичних варіантів, наприклад у генах *CYP2C19* чи *TPMT*, дозволяє прогнозувати індивідуальну чутливість пацієнта до певних ліків і зменшити ризик побічних ефектів. Такий підхід уже активно впроваджується у клінічну практику, зокрема в онкології та кардіології [19].

Таким чином, мутаційний аналіз, розглянутий з точки зору комп'ютерних наук, має фундаментальне значення для розробки інноваційних рішень у біоінформатиці та медицині. Метою є не тільки вдосконалення технологій аналізу, але й створення інтегрованих систем підтримки прийняття рішень, які допоможуть забезпечити ефективну діагностику та лікування захворювань.

Технології секвенування нового покоління (NGS) створили потужну основу для аналізу мутацій. Вони генерують величезні обсяги даних, які вимагають ефективного зберігання, обробки та аналізу за допомогою спеціалізованих програмних засобів.

Сучасні методи мутаційного аналізу спираються на інструменти

високопродуктивного секвенування, які генерують гігабайти даних за один цикл аналізу. Програмні платформи, такі як GATK (Genome Analysis Toolkit), ANNOVAR і SnpEff, забезпечують автоматизовану обробку цих даних. Зокрема, ANNOVAR використовується для функціональної анотації мутацій і визначення їх клінічного значення шляхом порівняння зі знаними базами даних, такими як ClinVar або COSMIC, де зберігаються відомості про клінічно значущі мутації [22, 23]. Інтеграція алгоритмів комп'ютерного зору та автоматизованих обчислень дозволяє не лише визначати місцезнаходження мутацій у геномі, а й оцінювати їх вплив на білкові структури та функції.

Одним із викликів мутаційного аналізу є ідентифікація "шуму" — псевдомутацій або технічних артефактів, що виникають під час секвенування. Комп'ютерні науки пропонують підходи до фільтрації таких артефактів, наприклад, за допомогою моделювання ймовірностей помилок на рівні секвенування або застосування алгоритмів типу DeepVariant, які використовують глибоке навчання для покращення точності виявлення варіантів [24].

Машинне навчання та ШІ дедалі частіше застосовуються для прогнозування клінічного значення мутацій. Наприклад, алгоритми, реалізовані у PolyPhen-2 або MutationTaster, використовують моделі ШІ для передбачення того, чи є певний генетичний варіант патогенним [25]. Інші підходи, такі як використання нейронних мереж, дозволяють автоматично класифікувати мутації залежно від їхнього функціонального впливу. Деякі системи також використовують мовні моделі, схожі на GPT, для анотації генетичних варіантів шляхом пошуку відповідних публікацій у науковій літературі.

Крім цього, аналіз великих когорт пацієнтів із подібними генетичними профілями став можливим завдяки технологіям розподілених обчислень, таких як Hadoop або Spark. Вони дозволяють обробляти петабайти даних, виконувати асоціативний аналіз геномів та виявляти рідкісні, але клінічно значущі мутації. Інструменти, такі як Galaxy, пропонують інтерактивні середовища для аналізу

мутацій, що робить технології більш доступними для медичних дослідників без спеціалізованих навичок програмування [26].

Результати мутаційного аналізу інтегруються у клінічні інформаційні системи для підтримки прийняття рішень. Це створює можливості для персоналізованої медицини, де вибір терапії базується на унікальному генетичному профілі пацієнта. Наприклад, застосування алгоритмів для прогнозування відповіді на імунотерапію може бути використано у лікуванні онкологічних захворювань. Наприклад, у випадку спадкових онкологічних захворювань, таких як рак молочної залози, аналіз мутацій у генах *BRCA1* та *BRCA2* дозволяє визначити ступінь ризику захворювання. На основі даних з NGS, сучасні алгоритми виявляють навіть рідкісні варіанти, що раніше залишалися поза увагою. Крім того, використання AI-моделей сприяє побудові прогнозів захворюваності та рекомендацій щодо профілактичних заходів, таких як генетичний моніторинг або профілактичне видалення тканин.

Окрім того, алгоритми кластеризації, такі як k-means або DBSCAN, застосовуються для аналізу гетерогенності пухлин. Завдяки цьому можна визначити домінуючі клональні мутації, що допомагає в адаптації терапії під специфічний профіль пухлини пацієнта. Наприклад, при раку легенів аналіз мутацій у генах *EGFR* або *ALK* дозволяє лікарям вибрати найбільш ефективну таргетну терапію, що значно підвищує виживаність пацієнтів [27, 28].

Комерційні платформи, такі як FoundationOne, пропонують інтегровані рішення, які поєднують секвенування з аналітичними інструментами для підтримки клінічних рішень. Крім аналізу, комп'ютерні системи забезпечують інтеграцію мутаційних даних у клінічні інформаційні системи. Платформи, такі як cBioPortal і OncoKB, дозволяють лікарям отримувати персоналізовані звіти для прийняття рішень, базуючись на молекулярному профілі пацієнта. Наприклад, використання cBioPortal у багатьох клініках дозволяє швидко інтегрувати дані секвенування в робочий процес онкологів, прискорюючи вибір терапії [5, 26].

Нарешті, комп'ютерні системи також сприяють вдосконаленню самих алгоритмів і підходів у мутаційному аналізі. Завдяки використанню систем з високою обчислювальною потужністю, таких як Google Cloud або Microsoft Azure, дослідники отримують можливість проводити великомасштабні обчислення, необхідні для оптимізації моделей прогнозування ризику.

Таблиця 1.1 наочно демонструє, як різні інструменти й технології сприяють кожному етапу мутаційного аналізу. Вона показує зв'язки між технічними рішеннями та їх клінічним значенням.

Таблиця 1.1 – Комп'ютерні інструменти та алгоритми для мутаційного аналізу в клінічній діагностиці

Етап аналізу	Інструменти/ Алгоритми	Функціонал	Приклад клінічного застосування
Обробка даних	FASTQC, Trimmomatic	Оцінка якості та очищення сирих даних секвенування	Видалення низькоякісних ридів для підвищення точності подальшого аналізу
Виявлення мутацій	GATK, DeepVariant	Ідентифікація варіантів у послідовностях ДНК	Виявлення мутацій, пов'язаних із раком грудей
Анотація мутацій	ANNOVAR, VEP	Інтерпретація функціональної ролі варіантів	Визначення клінічної значущості варіантів у генах BRCA1/BRCA2

Продовження таблиці 1.1

Прогнозування	PolyPhen-2, SIFT, MutationTaster	Оцінка впливу мутацій	Визначення патогенності нових варіантів
Інтеграція результатів	cBioPortal, OncoKB	Інтеграція даних у клінічні системи для підтримки прийняття рішень	Розробка персоналізованих схем лікування на основі мутацій
Оптимізація	Machine learning models (Random Forest, CNN)	Автоматизація та підвищення точності прогнозування	Побудова моделей для оцінки ризику розвитку хвороб

Таким чином, комп'ютерні науки виступають фундаментальною основою для розвитку мутаційного аналізу у клінічній діагностиці, роблячи його доступним, масштабованим та ефективним. Взаємодія обчислювальних технологій з біологічними знаннями дозволяє перетворювати складні генетичні дані у практичні рішення, що змінюють підходи до лікування та профілактики хвороб.

Виклики в аналізі мутацій із точки зору комп'ютерних наук охоплюють широкий спектр проблем, які варіюються від технічних аспектів опрацювання даних до інтерпретації результатів у клінічному контексті. Одним із ключових викликів є величезний обсяг даних, які генеруються сучасними технологіями секвенування, такими як платформи Illumina або Oxford Nanopore. Обробка таких даних потребує значних обчислювальних ресурсів, і навіть на сучасних суперкомп'ютерах час аналізу може бути суттєвим [28]. Це створює потребу в розробці алгоритмів, здатних працювати ефективно з великими обсягами інформації, зберігаючи високу точність.

Другий виклик полягає у складності вирівнювання ридів до референтного

геному. Алгоритми, такі як BWA та Bowtie, хоча й оптимізовані для вирішення цього завдання, все ще стикаються із проблемами в обробці геномів із високою кількістю повторів або структурними варіаціями [29]. Це може призводити до пропуску важливих мутацій, що має критичне значення в контексті клінічної діагностики, наприклад, у виявленні мутацій у генах, пов'язаних із раком, таких як TP53 або KRAS [30, 31].

Іншою серйозною проблемою є диференціація між справжніми мутаціями та артефактами секвенування. Артефакти можуть виникати внаслідок похибок у роботі секвенаторів або під час ПЛР-ампліфікації. Інструменти, як-от VarScan або MuTect2, розроблені для виявлення соматичних мутацій, намагаються розрізнити ці явища, але в складних випадках помилки залишаються можливими. Наприклад, для низькочастотних мутацій у пухлинних клітинах часто важко досягти високої точності, оскільки артефакти мають схожу частоту.

Розробка методів прогнозування функціонального впливу мутацій також залишається викликом. Алгоритми, такі як PolyPhen-2 та CADD, які використовують машинне навчання, покладаються на доступність якісних тренувальних наборів даних. Проте не всі варіанти мають добре охарактеризовані наслідки, особливо рідкісні або нові мутації. Це ускладнює прогнозування та часто вимагає ручного втручання для перевірки результатів.

Ще одна проблема пов'язана з інтеграцією біоінформатичних інструментів у клінічні робочі процеси. Дані секвенування та аналізу мутацій часто повинні відповідати строгим стандартам, таким як вимоги FDA чи європейських медичних агентств. Платформи, як-от BaseSpace від Illumina або Alissa от Agilent, розробляються для спрощення цієї інтеграції, але кожна клініка має свої унікальні вимоги, що створює додаткові труднощі [28, 32].

Нарешті, зберігання та управління даними є критичними питаннями. Системи, такі як Amazon Web Services або локальні рішення на основі Hadoop, пропонують масштабовані рішення для зберігання, але вимагають значних фінансових витрат і забезпечення відповідності стандартам конфіденційності,

таким як GDPR. Зокрема, секвенування персональних геномів пацієнтів вимагає впровадження спеціальних протоколів для захисту даних, що додає складності процесу.

На рисунку 1.4 запропоновано схему, яка відобразить основні етапи аналізу мутацій і місця, де виникають проблеми.



Рисунок 1.4 – Виклики у процесі аналізу мутацій

Таким чином, виклики в аналізі мутацій включають проблеми, які лежать на стику комп'ютерних наук, біології та медицини. Вирішення цих проблем вимагає мультидисциплінарного підходу, що поєднує сучасні досягнення в алгоритмах, обчислювальних системах та біологічних базах даних. Успішна

реалізація таких рішень може суттєво прискорити клінічний прогрес і впровадження персоналізованої медицини.

### 1.3 Аналіз методів підтримки діагностичних рішень у медичних інформаційних системах

Методи підтримки діагностичних рішень у медичних інформаційних системах спрямовані на покращення точності, швидкості та обґрунтованості клінічних висновків шляхом інтеграції сучасних технологій обробки даних та алгоритмів. Вони ґрунтуються на використанні передових методів аналізу інформації, таких як машинне навчання, моделі на основі правил, а також статистичні підходи, що дозволяють лікарям ефективніше інтерпретувати дані пацієнтів.

Системи підтримки прийняття діагностичних рішень (CDSS) є однією з найважливіших складових МІС. Вони базуються на принципі комбінування даних пацієнта з наявними клінічними знаннями для надання рекомендацій лікарям. Наприклад, такі системи можуть аналізувати результати лабораторних досліджень, візуалізацій та анамнезу для оцінки ймовірності певного діагнозу або виявлення потенційних ускладнень. Алгоритми, що використовуються у CDSS, можуть бути основані на машинному навчанні, де моделі навчаються на великих масивах даних для ідентифікації закономірностей, або на експертних правилах, що визначають стандартизовані дії залежно від введених параметрів.

Серед найбільш відомих прикладів таких систем можна назвати IBM Watson Health, яка використовує AI для обробки великих медичних даних, і Google Health, яка застосовує алгоритми глибокого навчання для діагностики хвороб на основі медичних зображень [13]. Наприклад, Watson може аналізувати медичні записи та надавати лікарям можливість швидше приймати рішення на

основі найсучасніших медичних знань. Інший приклад — це застосування алгоритмів глибокого навчання в обробці зображень, де технології, такі як Convolutional Neural Networks (CNN), використовуються для автоматичного виявлення пухлин або патологій на медичних зображеннях, таких як рентгенівські знімки або МРТ. Такий підхід значно знижує ймовірність людської помилки і дозволяє лікарям фокусуватися на більш складних випадках [15].

Крім алгоритмів машинного навчання, важливу роль відіграють статистичні методи. Наприклад, регресійний аналіз або кластеризація можуть бути використані для оцінки взаємозв'язку між різними клінічними параметрами пацієнтів та ймовірністю розвитку захворювань. Такі методи є особливо корисними для аналізу великих обсягів даних, які надходять з електронних медичних записів або систем моніторингу стану пацієнтів [7].

Ще одним аспектом є інтеграція баз знань у МІС. Такі бази знань можуть включати клінічні протоколи, наукові дослідження та рекомендації, що дозволяє CDSS надавати лікарям обґрунтовані поради на основі найактуальніших даних. Наприклад, система може запропонувати найбільш ефективний варіант лікування, ґрунтуючись на доказах, отриманих у масштабних клінічних випробуваннях. Інтеграція таких технологій в медичні інформаційні системи дозволяє здійснювати аналітику великих даних (big data analytics), що є ключовим інструментом для створення персоналізованих планів лікування. Медичні дані, які генеруються в результаті досліджень, тестів і лікарських спостережень, можуть бути зібрані і проаналізовані за допомогою методів data mining для виявлення прихованих закономірностей і тенденцій, що допомагає в прогнозуванні розвитку захворювань або визначенні найбільш ефективних методів лікування.

Незважаючи на величезний потенціал CDSS, їх інтеграція в існуючі медичні інформаційні системи викликає деякі проблеми. Основною проблемою є необхідність забезпечення точності та надійності моделей, які використовуються у таких системах. Це вимагає ретельної валідації алгоритмів і

навчальних даних, щоб уникнути помилкових результатів. Крім того, питання інтерпретованості моделей залишається актуальним, оскільки лікарі повинні розуміти, як саме система дійшла до того чи іншого висновку.

Іншою важливою характеристикою CDSS є можливість інтеграції з іншими медичними технологіями, такими як телемедицина та електронні медичні записи. Електронні медичні картки (EMR), що мають інтеграційні можливості з іншими медичними системами, що зберігають дані про пацієнтів, включаючи його історію хвороби, попередні діагнози, результати тестів та консультації лікарів [5, 12]. Завдяки цьому лікарі мають доступ до всієї необхідної інформації для прийняття обґрунтованого рішення, а системи підтримки прийняття рішень на основі цієї інформації можуть пропонувати рекомендації, що допомагають знижувати ризики помилок. Це дозволяє лікарям отримувати в реальному часі всю необхідну інформацію, а також покращує комунікацію між різними фахівцями, що працюють над лікуванням одного пацієнта. Більш того, деякі системи здатні забезпечити автоматичне оновлення клінічних протоколів на основі нових наукових даних або змін у медичних стандартах, що є важливим для адаптації до швидко змінюваного медичного середовища.

Використання таких технологій також несе важливі етичні питання. Наприклад, при створенні рекомендаційних систем у медицині потрібно враховувати не тільки точність алгоритмів, а й захист даних пацієнтів. Медичні інформаційні системи повинні відповідати міжнародним стандартам безпеки, таким як HIPAA в США або GDPR в Європі, щоб забезпечити конфіденційність особистої інформації пацієнтів. Зі зростанням обсягу медичних даних і розвитку технологій виникає потреба в оптимізації процесів для забезпечення високої ефективності при збереженні етичних стандартів. Крім того, необхідно враховувати, що навіть з найпотужнішими алгоритмами, лікарі повинні залишатися ключовими у прийнятті кінцевого рішення, адже технології є допоміжними, а не замінними для людського досвіду та знань.

Архітектура систем підтримки рішень, які використовують мутаційний

аналіз, є складним багаторівневим підходом, що інтегрує збирання, обробку та візуалізацію даних. Ці системи базуються на тісній взаємодії різних компонентів, таких як генетичні бази даних, аналітичні модулі й користувацькі інтерфейси, які забезпечують кінцевих користувачів точними й зрозумілими результатами (рис.1.5).



Рисунок 1.5 – Архітектура систем підтримки діагностичних рішень у МІС

На рисунку 1.5 вище зображено концептуальну схему, яка ілюструє ключові компоненти та їх взаємодію у системах підтримки діагностичних рішень. Вона демонструє, як об'єднання різноманітних джерел даних, аналітичних модулів та механізмів виведення інформації створює цілісну екосистему для прийняття рішень у медичній практиці.

Основою будь-якої такої системи є модуль збору даних, який об'єднує інформацію з різних джерел, включаючи генетичні аналізатори (наприклад, секвенатори), електронні медичні записи (EMR) і загальнодоступні бази даних, такі як ClinVar або COSMIC [22]. Ці джерела можуть включати як структуровані, так і неструктуровані дані, наприклад, сирі файли FASTQ або VCF з даними секвенування, медичні звіти та результати біомаркерного тестування. Інтеграція цих даних потребує не лише стандартизації форматів, але й попередньої обробки, яка включає нормалізацію, видалення помилок і первинний аналіз, наприклад, за допомогою ANNOVAR чи VEP (Variant Effect Predictor) [26].

Наступним ключовим етапом є аналітичний модуль, який об'єднує алгоритми для аналізу мутаційних даних. Він включає класифікацію варіантів, виявлення функціонально значущих змін та прогнозування їхнього впливу на клінічні рішення. Наприклад, система Watson for Genomics використовує машинне навчання для пошуку мутацій, які впливають на терапію онкологічних хворих, тоді як OncoKB інтегрує знання про біологічне значення мутацій з рекомендаціями щодо їх терапевтичного використання [23]. Аналіз часто доповнюється алгоритмами прогнозування, що враховують дані про експресію генів, білкові взаємодії й інші контекстуальні фактори.

Важливою частиною архітектури є модуль візуалізації, який забезпечує кінцевих користувачів зручним інтерфейсом для інтерпретації результатів. Наприклад, системи на кшталт cBioPortal дозволяють візуалізувати мутації в контексті клінічних даних, пропонуючи інтерактивні графіки та схеми, які легко зрозуміти навіть лікарям без ґрунтовних знань у геноміці. Для користувачів із більш технічними потребами системи надають доступ до сирих даних і

розширених інструментів аналізу.

Інтеграція всіх цих компонентів вимагає гнучкості в дизайні архітектури. Наприклад, використання API дозволяє системам взаємодіяти з іншими програмами, зокрема EMR, з мінімальними зусиллями. При цьому важливо враховувати регуляторні вимоги. У США, наприклад, FDA регулює використання програмного забезпечення, яке впливає на прийняття клінічних рішень. У Європі діє MDR, яка вимагає відповідності високим стандартам безпеки. У таких умовах дотримання регуляцій стає важливою частиною архітектурного дизайну, що забезпечує довіру користувачів і юридичну відповідність.

На практиці інтеграція таких систем демонструє значний потенціал. Наприклад, система Tempus Labs об'єднує генетичні дані, дані про терапії та клінічні результати, щоб надати лікарям персоналізовані рекомендації в режимі реального часу. Водночас, система Alissa Interpret працює на основі хмарних технологій, що дозволяє дослідникам і лікарям швидко проводити аналіз мутацій та ділитися даними між різними клінічними центрами [16].

Таким чином, архітектура систем підтримки рішень із використанням мутаційного аналізу демонструє приклад успішного поєднання передових обчислювальних технологій із потребами кінцевих користувачів. Завдяки інтеграції складних аналітичних модулів і зручних інтерфейсів, ці системи мають потенціал суттєво покращити якість діагностики й лікування.

#### 1.4 Постановка задачі дослідження

Постановка задачі дослідження є важливим етапом у розробці та реалізації будь-якого наукового проєкту, адже вона визначає ключові напрями роботи та обмеження, необхідні для досягнення результатів. У контексті аналізу методів

підтримки діагностичних рішень на основі мутаційного аналізу, актуальність теми обумовлена стрімким розвитком персоналізованої медицини, яка базується на індивідуальних геномних особливостях пацієнтів.

Існуючі методи підтримки рішень діагностування використовують класичні алгоритми аналізу даних, машинне навчання та статистичні моделі для виявлення патологій на основі клінічної інформації, зображень або генетичних даних. Однак ці методи мають такі обмеження, як недостатня точність у випадках рідкісних або комплексних мутацій, висока залежність від якості вхідних даних та труднощі інтеграції різнорідних джерел інформації, що не дає можливості забезпечити підтримку діагностичних рішень в умовах великих обсягів даних та різноманітності генетичних змін.

Для вирішення цієї задачі з високою точністю може бути використаний підхід на основі мутаційного аналізу, що передбачає ідентифікацію патогенних мутацій за допомогою сучасних алгоритмів обробки секвенувальних даних, таких як CRISPR-баркодинг і технології NGS. Такий підхід дозволяє не лише точно визначати зміни в геномі, але й оцінювати їхній вплив на розвиток захворювань, використовуючи системи прогностичного аналізу ризиків і алгоритми машинного навчання для інтеграції результатів у клінічну практику.

Об'єктом дослідження є процес діагностування на основі аналізу МІС. Предметом дослідження є методи підтримки діагностичних рішень МІС.

Сучасні медичні інформаційні системи мають забезпечити лікарів інструментами для швидкої й точної інтерпретації генетичних даних пацієнта, але існують численні проблеми, пов'язані з обробкою великих обсягів даних, інтеграцією алгоритмів машинного навчання і створенням інтероперабельних платформ. Водночас складність таких систем вимагає глибокого аналізу існуючих методів і підходів для розробки рішень, які б поєднували точність, ефективність і зрозумілість для користувача.

Основною метою цього дослідження є створення концептуальної моделі для підтримки прийняття рішень у медичних інформаційних системах, яка

використовує мутаційний аналіз для діагностики та прогнозування захворювань. Для цього необхідно досягти кількох завдань:

- провести аналіз предметної області для визначення основних проблем і напрямів розвитку медичних інформаційних систем із підтримкою мутаційного аналізу;

- визначити проблеми та недоліки існуючих методів обробки мутаційних даних та їх інтеграції у діагностичні процеси, особливо в контексті систем підтримки прийняття рішень;

- розробити підхід, який забезпечуватиме точну інтерпретацію мутаційних ознак із урахуванням клінічного контексту;

- удосконалити метод, здатного інтегрувати геномні дані в реальний клінічний процес, водночас зберігаючи простоту використання для медичного персоналу;

- провести тестування запропонованих рішень на експериментальних або симуляційних даних для оцінки їхньої ефективності.

Це дослідження також передбачає порівняльний аналіз сучасних підходів у діагностуванні з використанням мутаційного аналізу.

Важливим аспектом є створення моделі, яка не тільки обробляє мутаційні дані, але й враховує інші клінічні параметри, дозволяючи інтегрувати різноманітні джерела інформації в єдину систему підтримки рішень.

У рамках даного дослідження ставиться задача розробки методів підтримки діагностичних рішень у МІС, що ґрунтуються на аналізі мутаційних ознак для покращення точності та ефективності діагностики захворювань, зокрема онкологічних. Метою є удосконалення існуючих підходів для підтримки клінічних рішень, а також розробка методів, що можуть враховувати різноманітні генетичні маркери для персоналізованої медицини.

Завданням є створення механізмів для автоматизованого аналізу мутацій у контексті діагностики на основі даних, що включають секвенування геному, лабораторні результати та інші клінічні відомості. Важливою частиною

дослідження є розробка таких методів, які можуть бути інтегровані у вже існуючі медичні інформаційні системи та підтримувати процес прийняття рішень на всіх етапах лікування — від виявлення захворювання до вибору оптимальних терапевтичних стратегій. Для цього необхідно зібрати, обробити та інтегрувати великі обсяги генетичних і клінічних даних, що дозволяють підвищити точність прогнозування і адаптувати лікування до індивідуальних потреб пацієнта.

У рамках цього дослідження також передбачається вивчення питання забезпечення сумісності між різними платформами МІС, а також покращення їхнього функціоналу шляхом інтеграції штучного інтелекту та машинного навчання для автоматизації інтерпретації генетичних даних. Необхідно врахувати технічні, етичні та юридичні аспекти, зокрема захист персональних даних пацієнтів, а також створення алгоритмів, що можуть враховувати як загальні медичні протоколи, так і індивідуальні особливості кожного пацієнта.

Дослідження також охоплює аналіз існуючих методів підтримки діагностичних рішень у медичних інформаційних системах, а також визначення можливих шляхів удосконалення існуючих алгоритмів для підвищення точності діагностики та прогнозування.

Очікується, що результати роботи сприятимуть розвитку методів діагностики у персоналізованій медицині, забезпечуючи лікарів інструментами для ухвалення точніших та швидших рішень. Це дозволить покращити якість надання медичних послуг, знизити ризики помилкових діагнозів і оптимізувати використання ресурсів у сфері охорони здоров'я.

## **2 ПІДТРИМКА ДІАГНОСТИЧНИХ РІШЕНЬ У МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ З ВИКОРИСТАННЯМ МУТАЦІЙНОГО АНАЛІЗУ**

2.1 Розробка підходу до підтримки діагностичних рішень на основі мутаційного аналізу

Розробка підходу до підтримки діагностичних рішень на основі мутаційного аналізу є складним, але перспективним напрямом у біоінформатиці та медичних інформаційних системах. У сучасній медичній практиці аналіз генетичних даних пацієнтів дозволяє не лише ідентифікувати причини патологій, але й передбачати розвиток захворювань і обирати оптимальні методи лікування. Цей процес вимагає комплексного підходу, який включає автоматизацію збору даних, їх обробку, аналіз і інтеграцію результатів у клінічний контекст.

Основним етапом розробки такого підходу є створення алгоритмів для інтерпретації генетичних варіантів. Наприклад, застосування методів машинного навчання дає змогу автоматично класифікувати мутації за їх клінічним значенням, що є критично важливим для скорочення часу, необхідного для прийняття рішень. Розвиток технологій NGS дозволив отримувати величезні обсяги генетичних даних, що потребують ефективних підходів до їх аналізу. Автоматизація цього процесу стала можливою завдяки розробці інструментів, таких як Somatic, які здатні проводити швидкий аналіз соматичних мутацій і визначати їхній зв'язок із клінічними фенотипами.

Наступним важливим елементом є адаптація отриманих даних до конкретних клінічних завдань. Це включає створення моделей прогнозування, які враховують як генетичні, так і інші клінічні параметри пацієнта. Застосування штучного інтелекту та великих даних у цій сфері дозволяє формувати прогностичні моделі, що враховують комплексну взаємодію між генетичними

варіантами та їх впливом на розвиток захворювань. Наприклад, обробка даних з мутаційного аналізу може вказати на те, які пацієнти матимуть підвищений ризик ускладнень або яка терапія буде для них найбільш ефективною.

Інтеграція результатів аналізу у медичні інформаційні системи є ще одним ключовим завданням, яке потребує розробки стандартизованих підходів до обміну даними. У цьому контексті важливим є використання електронних медичних записів (EHR), які забезпечують можливість зберігання, аналізу та використання генетичних даних у режимі реального часу. Наприклад, інтеграція модуля, який дозволяє швидко отримати результати мутаційного аналізу та врахувати їх у клінічному рішенні, може значно підвищити точність діагностики та ефективність лікування [37].

Розробка таких підходів вимагає співпраці між біоінформатиками, лікарями, інженерами та фахівцями з інформаційних технологій, як показано на рисунку 2.1.



Рисунок 2.1 – Ролі в процесі висування вимог до процесу розробки систем підтримки діагностичних рішень.

Лише за умов мультидисциплінарного підходу можливо створити інструменти, які враховують потреби клінічної практики та наукової точності. Розвиток цієї сфери також вимагає фінансування досліджень, побудови спеціалізованих обчислювальних платформ і забезпечення доступу до актуальних даних про генетичні варіанти.

SComatic — це інструмент, що спеціалізується на аналізі соматичних мутацій і дозволяє швидко та ефективно виявляти генетичні варіанти у зразках пацієнтів. Його основною перевагою є здатність інтегрувати результати аналізу з іншими даними в межах MIC. SComatic забезпечує можливість ідентифікувати клінічно значущі мутації та формувати висновки, які можуть використовуватися для підтримки діагностичних рішень.

SComatic є унікальним інструментом, розробленим для високоточного аналізу соматичних мутацій, що робить його ключовим елементом у підтримці діагностичних рішень [33]. Його основна функція полягає у швидкій та точній ідентифікації генетичних варіантів у даних секвенування, які мають значення для клінічної практики. З точки зору комп'ютерних наук, SComatic демонструє низку важливих аспектів, які можуть бути використані для покращення інформатики, алгоритмів аналізу даних, а також інтеграції в складні медичні інформаційні системи.

Аналіз соматичних мутацій із використанням комп'ютерних алгоритмів є складним і важливим процесом, що включає як обробку великих обсягів даних, так і їхнє інтелектуальне осмислення. Програма, яка аналізує такі дані, побудована на принципах обчислювальної біології, біоінформатики та прикладної статистики. Соматичні мутації, які виникають протягом життя людини в певних клітинах, відіграють ключову роль у вивченні таких захворювань, як рак. Використання сучасних технологій геномного секвенування дозволяє отримати детальну інформацію про ці мутації, а відповідні алгоритми забезпечують їх ефективний аналіз.

З точки зору архітектури, програма починається з обробки вхідних даних.

Це можуть бути файли у форматах VCF (Variant Call Format) чи BAM (Binary Alignment/Map), які є стандартами в біоінформатиці. Наприклад, VCF-файл містить інформацію про типи мутацій (заміни нуклеотидів, інсерції чи делеції), позиції у геномі та метрики якості, такі як частота алеля чи покриття. На початковому етапі програма виконує парсинг файлів, застосовуючи бібліотеки на кшталт HTSeq або PyVCF, щоб структурувати та нормалізувати дані [34].

Однією з важливих складових є використання "панелі нормалей" (Panel of Normals, PON), яка дозволяє фільтрувати мутації, притаманні нормальним клітинам, та зосереджуватись на патологічних. З технічного боку, це включає порівняння вхідних даних із попередньо збереженими масивами мутацій із нормальних зразків, що вимагає оптимізованих алгоритмів пошуку. Наприклад, бібліотека NumPy у Python або навіть бази даних SQLite можуть використовуватись для швидкого пошуку збігів.

Додатково, програма виконує фільтрацію даних за заданими параметрами, такими як якість секвенування або наявність мутацій у певних клітинних типах. Тут застосовуються методи обробки сигналів, а також статистичні моделі, зокрема байєсові оцінки для визначення ймовірності функціональної значущості мутацій. Такі пакети, як SciPy або Pandas, дозволяють реалізувати складні обчислення та забезпечити масштабованість для великих наборів даних.

На етапі виведення результатів програма створює фільтровані списки мутацій, збагачені додатковою інформацією. Наприклад, функція ANNOVAR може використовуватись для анотації мутацій із зазначенням контексту їх розташування в геномі (упорядковуючі чи регуляторні області). Ці результати також інтегруються з базами даних, такими як COSMIC (Catalogue of Somatic Mutations in Cancer) або з такими базами знань як NCBI, що дозволяє співставляти мутації з відомими патогенними варіантами [35].

З погляду практичної значущості, результати програми мають великий потенціал у клінічних дослідженнях. Наприклад, аналіз частот мутацій у різних типах клітин дозволяє зрозуміти специфіку патогенезу захворювань. Якщо

мутація часто зустрічається в епітеліальних клітинах, але рідко — у мієлоїдних, це може свідчити про її тканинну специфічність. Такі програми вже знаходять застосування, зокрема у компаніях, таких як Illumina чи Tempus, для розробки персоналізованих підходів до лікування.

Аналіз соматичних мутацій з точки зору комп'ютерних наук є втіленням багаторівневого процесу, що охоплює обробку даних, алгоритмічний аналіз та аналітичне моделювання. Програма, яка здійснює такий аналіз, розпочинається з етапу читання вхідних даних, де застосовуються інструменти для роботи з форматами файлів VCF або BAM, які є стандартом у галузі біоінформатики [36]. Для цього можуть використовуватись бібліотеки на кшталт PyVCF, що дозволяє структурувати дані для подальшого використання. Дані з таких файлів включають інформацію про тип мутацій, їх позиції в геномі та показники якості, такі як частота алеля чи глибина покриття [37].

Після цього здійснюється фільтрація та контроль якості даних. Алгоритми аналізу перевіряють мутації на відповідність пороговим значенням, наприклад, мінімальній глибині зчитування або відсутності систематичних помилок, таких як зміщення ланцюгів (strand bias). У цьому процесі важливим елементом є використання "панелі нормалей" (Panel of Normals, PON), яка дозволяє виключити мутації, характерні для нормальних клітин. Така фільтрація здійснюється через порівняння вхідних даних із попередньо створеним референтним списком мутацій, використовуючи алгоритми оптимізації пошуку, наприклад, індексування на основі хеш-таблиць або бази даних SQLite.

На етапі аналізу мутацій програма розраховує показники, які допомагають у їх інтерпретації. Зокрема, обчислюється частота варіантного алеля (variant allele frequency, VAF), яка вказує, яка частка зчитувань відповідає мутації, а також статистично значущі характеристики, включаючи оцінки якості та зміщення ланцюгів. Для цього часто використовуються бібліотеки обчислювальної статистики, як-от SciPy, і математичні моделі, зокрема баєсові підходи.

Ключовим компонентом є анотація мутацій. На цьому етапі програма збагачує дані додатковою інформацією про геномний контекст, наприклад, визначаючи, чи є мутація в кодуєчій області, в регуляторній зоні або поблизу сайтів сплайсингу. Одним із популярних інструментів для таких задач є ANNOVAR, що дозволяє інтегрувати дані з баз даних, таких як dbSNP або ClinVar, для визначення клінічної значущості мутацій [38].

Результати програми виводяться у вигляді відфільтрованих і анотованих даних, а також можуть включати статистичні підсумки й візуалізації. Наприклад, розподіл мутацій за клітинними типами чи розподіл частоти алеля можна візуалізувати за допомогою бібліотек Matplotlib або Seaborn. Такі візуалізації допомагають ідентифікувати закономірності, які були б важко помітні у числових таблицях.

Практичне застосування результатів такого аналізу є широким. У дослідженнях раку виявлення ключових мутацій, що сприяють росту пухлини, дозволяє зрозуміти молекулярні механізми захворювання. У сфері персоналізованої медицини аналіз мутацій допомагає створювати індивідуальні плани лікування, адаптовані до генетичного профілю пацієнта. Наприклад, мутації в гені EGFR, виявлені за допомогою таких програм, можуть вказувати на чутливість пацієнта до інгібіторів тирозинкінази. Крім того, ці інструменти використовуються для відкриття біомаркерів, що дозволяють діагностувати або прогнозувати перебіг захворювання, наприклад, за допомогою мутацій, зареєстрованих у базі COSMIC.

Одна з ключових переваг SComatic — це його здатність працювати з великими обсягами даних геномного секвенування. Сучасні алгоритми, які використовуються в інструменті, дозволяють обробляти дані у форматах VCF або BAM, що відповідають стандартам у галузі біоінформатики. SComatic використовує адаптивні методи фільтрації та класифікації мутацій, забезпечуючи мінімізацію хибнопозитивних результатів. Це особливо важливо в контексті клінічного застосування, де навіть невелика помилка може призвести

до некоректного діагнозу.

Інтеграція SComatic із зовнішніми базами даних, такими як COSMIC, ClinVar і OncoKB, дозволяє автоматизувати процес визначення клінічної значущості мутацій. З комп'ютерної точки зору, це демонструє важливість використання API та систем обробки запитів для доступу до актуальних даних. Наприклад, OncoKB надає структурувану інформацію про онкогенетичні варіанти, а ClinVar забезпечує доступ до інформації про клінічну патологію мутацій [39]. Такі інтеграції дозволяють створювати зручні й ефективні робочі процеси для користувачів, що є ключовим аспектом для комп'ютерних систем підтримки рішень.

Окрім аналізу мутацій, SComatic підтримує створення звітів у форматі, що легко інтегрується з МІС. Це досягається завдяки використанню стандартних форматів даних, таких як HL7 або FHIR, які широко використовуються в системах електронних медичних записів (EHR). З точки зору програмного забезпечення, це вимагає використання протоколів взаємодії, забезпечення конфіденційності даних, а також відповідності регуляторним стандартам, як-от GDPR або HIPAA.

SComatic також має потенціал для використання у створенні прогностичних моделей, які враховують як генетичні, так і клінічні дані. Наприклад, інтеграція з інструментами машинного навчання, такими як Scikit-learn або TensorFlow, дозволяє розробляти алгоритми, що передбачають ризик прогресування захворювання або відповіді на терапію. Подібні системи вже знаходять застосування, наприклад, у Tempus Labs, де використовуються аналітичні модулі для створення персоналізованих рекомендацій на основі геномних даних.

Нарешті, важливо зазначити, що SComatic є відкритою платформою, яка дозволяє налаштовувати робочі процеси під специфічні потреби користувачів. Це робить його гнучким інструментом, здатним адаптуватися до різних сценаріїв клінічного застосування. Наприклад, у клінічних дослідженнях, де використовуються високоточні дані секвенування CRISPR-експериментів,

SComatic може бути адаптований для виявлення мутацій, пов'язаних із специфічними генетичними маніпуляціями.

Таким чином, роль SComatic виходить за рамки простого інструменту аналізу мутацій, демонструючи широкий спектр застосувань у медицині, біоінформатиці та комп'ютерних науках. Його здатність інтегрувати дані, автоматизувати процеси і створювати зручні робочі середовища підкреслює його значення як частини систем підтримки рішень у сучасній персоналізованій медицині.

Розв'язання задачі, поставленої в першому розділі, передбачає створення фреймворку, який об'єднує аналіз соматичних мутацій, оцінку їх клінічної значущості, прогнозування ризиків і планування терапії. В основі цього підходу лежить поетапний процес, який враховує як наявні стандартизовані інструменти, так і новаторські рішення, що потребують розробки чи удосконалення. Такий підхід забезпечить системний і практичний внесок у розвиток персоналізованої медицини.

На першому етапі необхідно інтегрувати сирі дані геномного секвенування з наявних форматів (наприклад, VCF, BAM, або FASTQ) у єдину оброблювану структуру. Для цього буде використано існуючі інструменти попередньої обробки, такі як ANNOVAR для анотації варіантів і GATK для калібрування варіантів. Цей крок забезпечує стандартизацію даних та створює основу для подальшого аналізу. Особливу увагу слід приділити перевірці наявності артефактів, які можуть спотворити результати, що особливо важливо у клінічному контексті.

Другий етап передбачає застосування SComatic для аналізу соматичних мутацій. Основна задача на цьому етапі полягає у виявленні клінічно значущих мутацій із використанням панелей нормалей (PoN), узяті з інформації на основі баз даних, таких як OncoKB і COSMIC, для ідентифікації функціональних змін у генах. Удосконалення цього етапу може включати розширення можливостей SComatic для інтеграції з локальними клінічними базами даних і налаштування

параметрів аналізу для специфічних потреб користувачів (знову-ж-таки, створені під спеціальні потреби панелі нормалей) [33].

Наступний крок полягає в розробці або адаптації модуля для оцінки ризиків, який базується на мультифакторному аналізі генетичних і клінічних даних. Прогностична модель буде використовувати машинне навчання для врахування широкого спектра характеристик пацієнта. Наприклад, система Tempus Labs вже успішно застосовує подібний підхід для створення персоналізованих профілів ризиків, що дозволяє прогнозувати ймовірність рецидивів або прогресування захворювання.

Узагальнений процес обробки файлів показано на рисунку 2.2:



Рисунок 2.2 – Етапи мутаційного аналізу

Кінцевий етап — це інтеграція всіх попередніх компонентів у медичні інформаційні системи (МІС). Основним завданням цього етапу є забезпечення безперешкодної передачі даних і можливості їхнього використання у клінічній

практиці. Інтеграція повинна відповідати регуляторним стандартам, таким як GDPR у Європі чи HIPAA у США, що забезпечить конфіденційність і безпеку даних. Прикладом успішної інтеграції є система Epic, яка дозволяє лікарям отримувати генетичні дані пацієнтів у зрозумілому форматі безпосередньо в електронних медичних записах.

Побудова фреймворку на основі SComatic є актуальною через кілька причин. По-перше, цей підхід дозволяє адаптувати існуючі алгоритми до специфічних потреб клінічної практики. По-друге, він сприяє розвитку персоналізованої медицини через створення інструментів, які враховують унікальні генетичні профілі пацієнтів.

Архітектура запропонованого фреймворку на основі SComatic є складним багаторівневим підходом, який інтегрує сучасні методи аналізу мутаційних даних, алгоритмічні оптимізації та можливості для масштабування. Розробка такого фреймворку враховує потреби в точності, продуктивності та інтерактивності, що є критично важливими для підтримки діагностичних рішень на основі генетичного аналізу.

На базовому рівні архітектура складається з трьох основних компонентів: рівня обробки даних, рівня аналітики та рівня візуалізації. На рівні обробки даних фреймворк забезпечує інтеграцію та попередню обробку великих обсягів мутаційних даних, включаючи зчитування вхідних файлів (наприклад, VCF або BED форматів), фільтрацію шуму за допомогою баз даних панелей нормальних варіантів, таких як PON, та забезпечення стандартизації даних за допомогою існуючих інструментів на кшталт GATK і VCFtools. Для цього рівня важливим є забезпечення гнучкості обробки, яка може бути реалізована через контейнеризовані сервіси, такі як Docker, що дозволяють масштабувати роботу з масивами даних.

Рівень аналітики в цьому фреймворку є основним механізмом обчислень. Тут застосовуються алгоритми для ідентифікації мутацій з високим впливом. Наприклад, можна використовувати методи машинного навчання, як-от

XGBoost, для класифікації соматичних мутацій на основі багатьох параметрів, включаючи частоту алелей та їх розподіл по клітинах. Додатково, цей рівень інтегрує ресурси, такі як бази даних ClinVar, COSMIC або dbSNP, для анотації мутацій. Ця анотація дозволяє визначити функціональну значущість кожної мутації, наприклад, вплив на білкові домени або участь у відомих сигнальних шляхах. Аналітичний модуль також відповідає за обчислення статистичних метрик і генерацію прогнозів, які можуть використовуватися лікарями для прийняття рішень.

Візуалізація є ключовою для забезпечення зрозумілості результатів. Фреймворк передбачає динамічну генерацію графіків і інтерактивних таблиць через інструменти на кшталт Plotly або ggplot2. Наприклад, важливим завданням є візуалізація розподілу мутацій за різними типами клітин, що може вказувати на тканинну специфічність. Інтерактивні панелі, створені за допомогою R Shiny або Python Dash, дозволяють користувачам змінювати параметри фільтрації та аналізу в реальному часі, що значно підвищує ефективність роботи з системою.

Ключовим аспектом архітектури є її модульність. Всі компоненти фреймворку побудовані так, щоб вони могли інтегруватися з іншими інструментами або оновлюватися незалежно. Наприклад, модуль анотації може бути оновлений для роботи з новими версіями баз даних, такими як оновлений COSMIC, без зміни інших частин системи. Більш того, архітектура може підтримувати розподілену обробку за допомогою платформ, як-от Apache Spark, для оптимізації аналізу великих даних.

Окрім цього, ми ставимо перед собою задачу включити до фреймворку можливість планування терапії за рахунок підбору гРНК.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) — це революційна технологія редагування геному, яка дозволяє вченим точно змінювати ДНК в живих організмах. В її основі лежить природний механізм захисту бактерій від вірусів. Бактерії зберігають фрагменти вірусної ДНК у спеціальних ділянках геному (CRISPR-локусах) і використовують їх разом із

білком Cas (найчастіше Cas9) для ідентифікації та знищення вірусів у разі повторного зараження. Завдяки цьому механізму, вчені змогли створити універсальний інструмент для роботи з геномами різних організмів [40].

У терапії CRISPR використовується для лікування генетичних захворювань шляхом виправлення мутацій у ДНК. Наприклад, технологія вже показала перспективи у лікуванні серповидно-клітинної анемії, муковісцидозу та деяких форм онкологічних захворювань. Замість того щоб лише полегшувати симптоми хвороби, CRISPR дозволяє усувати її причину на молекулярному рівні, що відкриває нові горизонти в медицині. Особливо важливо, що ця технологія дозволяє не лише видаляти шкідливі мутації, але й вводити корисні генетичні зміни [41].

Ключовим компонентом CRISPR-системи є гРНК (гайдова РНК, або guide RNA), яка слугує навігатором для білка Cas9. Вона визначає, де саме відбудеться редагування, зв'язуючись із відповідною ділянкою ДНК через принцип комплементарності. З огляду на це, правильний вибір гРНК є критично важливим. Неправильно підібрана гРНК може призвести до «позамішеневих» змін, які потенційно є небезпечними та можуть викликати небажані побічні ефекти. Наприклад, мутації в нецільових ділянках ДНК можуть вплинути на життєво важливі гени або призвести до нестабільності геному.

Процес підбору гРНК базується на біоінформатичних алгоритмах, які аналізують геном, ідентифікують цільові послідовності та прогнозують можливість позамішеневих ефектів. Важливість цього завдання обумовлена високою складністю взаємодії між гРНК і ДНК. Ефективність редагування залежить від багатьох факторів, таких як довжина гРНК, її термодинамічна стабільність та відсутність значущої схожості з нецільовими ділянками. Таким чином, вміння точно підбирати гРНК є фундаментальним етапом у розробці CRISPR-основних терапій, які забезпечують безпеку та ефективність лікування.

Планування терапії на основі CRISPR-технологій є більш інноваційним і менш усталеним етапом. Тут важливо забезпечити зручність використання

інструментів для лікарів, які не є експертами у CRISPR. Використання базових рекомендацій на основі баз даних, таких як CRISPR Therapeutics, дозволить автоматизувати частину процесу. Наприклад, якщо пацієнт має соматичну мутацію у гені KRAS, модуль може рекомендувати відповідні терапевтичні стратегії або клінічні дослідження [41, 42].

Тож, за допомогою підключення до відповідних баз даних після виконання базового методу, допоможе розширити функціонал та надати можливість кінцевим користувачам-лікарям ефективніше інтерпретувати результати виконання алгоритму.

Загалом, запропонована архітектура орієнтована на інтеграцію існуючих потужних інструментів з адаптованими алгоритмами для специфічних задач. Це дозволяє створити фреймворк, який є як технічно складним, так і зручним для кінцевого користувача, що відповідає сучасним вимогам до систем підтримки діагностичних рішень.

Елементи фреймворку включають в себе:

–Mutation-Based Diagnostic Tool — модуль для автоматичного визначення мутацій, їхньої клінічної значущості та впливу на діагноз. Один із важливих компонентів архітектури — це модуль для завантаження та попередньої обробки мутаційних даних. Це включає фільтрацію вихідних файлів, таких як VCF або BAM, для видалення шумових мутацій і підтвердження якості даних. Далі, за допомогою спеціальних алгоритмів, можна здійснити анотацію мутацій, що дозволяє визначити їх клінічну значущість і можливі біологічні наслідки. Цей етап може бути реалізований за допомогою таких інструментів, як ANNOVAR або VEP (Variant Effect Predictor);

–Prognostic Risk Assessment — система для оцінки ризиків захворювання на основі генетичних і клінічних даних. Другим важливим компонентом є система прогнозування ризиків, яка використовує як генетичні, так і клінічні дані для оцінки ймовірності розвитку захворювань. Для цієї частини можна застосувати методи машинного навчання, наприклад, алгоритми класифікації,

що дозволяють передбачити, які мутації можуть мати найбільший вплив на розвиток хвороби. Використання таких бібліотек, як scikit-learn у Python, дозволяє швидко протестувати різні моделі, наприклад, логістичну регресію чи дерева рішень, для оцінки медичних ризиків;

–CRISPR-Driven Precision Therapy Planner — інструмент, що використовує результати мутаційного аналізу для планування терапій, зокрема CRISPR-технологій. Для підтримки персоналізованого підходу в лікуванні можна додати інструмент, який враховує індивідуальні мутаційні профілі пацієнтів і використовує результати мутаційного аналізу для планування терапії, включаючи використання CRISPR-технологій. Цей компонент може бути інтегрований з існуючими базами даних, що містять інформацію про терапевтичні таргети та клінічні дослідження;

–Integration into an Electronic Health Record (EHR) — модуль, який забезпечує інтеграцію з електронними медичними записами для зберігання, обробки та використання даних у клінічній практиці. Інтеграція з електронними медичними записами (EHR) є ще одним важливим аспектом архітектури. Це дозволяє зберігати всі дані про пацієнта в єдиній системі, що забезпечує зручний доступ до інформації для лікарів та медичних працівників, а також дозволяє автоматично оновлювати медичні карти з новими результатами аналізів.

Актуальність такого підходу зумовлена зростанням обсягу генетичних даних, які потребують швидкої та точної інтерпретації, а також потребою у більшій автоматизації процесів у клінічній практиці. Це дозволяє зменшити кількість помилок, підвищити ефективність лікування і зробити медичну допомогу доступнішою.

Підсумовуючи, підхід до підтримки діагностичних рішень на основі мутаційного аналізу повинен бути системним і враховувати як наукові, так і практичні аспекти. Інструменти на кшталт SComatic є важливим етапом на цьому шляху, адже вони спрощують аналіз даних, дозволяючи лікарям приймати більш обґрунтовані рішення. Окрім того, інтеграція таких інструментів у медичні

інформаційні системи відкриває нові можливості для персоналізованої медицини, роблячи її більш ефективною та доступною.

## 2.2 Розробка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу

Розробка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу є важливим завданням для створення ефективних інструментів підтримки діагностичних рішень у сучасних медичних інформаційних системах. Такий підхід дозволяє інтегрувати отримані дані з генетичного аналізу та враховувати як клінічні, так і індивідуальні характеристики пацієнта при прийнятті рішень.

Одним із ключових аспектів цієї розробки є створення алгоритмів, здатних виявляти найбільш релевантні варіанти мутацій з великого обсягу даних. Наприклад, інструменти типу SComatic дозволяють класифікувати соматичні мутації за їх значущістю, визначаючи, які саме з них мають критичний вплив на розвиток захворювань. Цей підхід базується на використанні великих даних і алгоритмів машинного навчання, що аналізують зв'язок між мутаціями та клінічними фенотипами. Така автоматизація знижує ризик людської помилки і дозволяє швидше отримувати результати, які можуть стати основою для прийняття рішень.

Важливим етапом є врахування обмежень, що виникають у процесі інтерпретації мутацій. Наприклад, для формування рекомендацій щодо лікування необхідно враховувати не лише генетичні дані, але й такі фактори, як вік, стать, наявність супутніх захворювань, а також можливі протипоказання до терапії. Алгоритми повинні враховувати ці обмеження шляхом адаптації отриманих даних до конкретного пацієнта, що є основою персоналізованого

підходу в медицині.

Одним із перспективних методів є використання систем багатокритеріальної оптимізації, які дозволяють формувати набір рішень з урахуванням усіх релевантних факторів. У цьому контексті важливою є концепція домінуючих рішень, яка передбачає вибір оптимальних варіантів лікування, які мають максимальну ефективність і мінімальні ризики для пацієнта. Такі рішення можуть включати рекомендації щодо терапії, прогностичні оцінки ризиків і можливість участі у клінічних дослідженнях.

Крім того, інтеграція отриманих результатів у медичні інформаційні системи потребує розробки зручних і зрозумілих інтерфейсів, які дозволяють лікарям швидко отримувати доступ до релевантної інформації. Наприклад, формування візуалізацій, що відображають зв'язок між мутаціями та їх впливом на терапевтичні результати, може значно спростити процес прийняття рішень. Це також сприяє підвищенню довіри до автоматизованих систем і стимулює їх широке впровадження в клінічну практику.

Загалом, розробка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу є багаторівневим процесом, який поєднує біоінформатику, медичну практику та інформаційні технології. Впровадження таких підходів дозволяє покращити якість діагностики, підвищити ефективність лікування та знизити витрати на медичні послуги, що є ключовим завданням сучасної медицини. Інструменти на основі таких систем можуть стати невід'ємною частиною майбутніх медичних інформаційних платформ.

### **3 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДТРИМКИ ДІАГНОСТИЧНИХ РІШЕНЬ У МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ НА ОСНОВІ МУТАЦІЙНОГО АНАЛІЗУ**

#### **3.1 Концепція та вимоги до розробки технології**

У сучасній медицині інформаційні технології відіграють ключову роль у покращенні точності діагностики, ефективності лікування та організації персоналізованої допомоги пацієнтам. Концепція розробки технології підтримки діагностичних рішень у медичних інформаційних системах базується на інтеграції багатовимірних даних, таких як геномні, клінічні та демографічні, що дозволяє забезпечити новий рівень точності в аналізі хвороб та прогнозуванні їх перебігу. Аналіз мутаційного профілю є важливою складовою цього підходу, оскільки саме мутації часто є ключовими маркерами захворювань, зокрема онкологічних.

Розробка таких технологій має враховувати необхідність автоматизації аналізу даних, адже ручна обробка є не лише трудомісткою, але й пов'язаною із ризиком людської помилки. У цьому контексті стає важливою розробка алгоритмів, здатних у реальному часі обробляти великі обсяги даних, виявляти мутації та формувати на їх основі рекомендації. Крім того, сучасні медичні системи мають бути інтегровані із електронними медичними записами, щоб забезпечити безперервний доступ до інформації та її ефективне використання в клінічній практиці.

SComatic, як інструмент для аналізу соматичних мутацій, є надзвичайно актуальним у цьому контексті. Він дозволяє автоматизувати процес інтерпретації мутацій, аналізувати їх вплив на біологічні функції, а також враховувати взаємодії мутацій у рамках генетичних мереж. Це робить його незамінним інструментом для формування діагностичних гіпотез, прогнозування ризиків та планування персоналізованої терапії. Застосування SComatic дозволяє

забезпечити як структурований підхід до аналізу, так і гнучкість у адаптації під різні клінічні потреби.

Однією з ключових вимог до розробки таких технологій є їхня здатність інтегрувати дані з різних джерел, включаючи секвенування нового покоління (NGS), аналіз біомаркерів і клінічну інформацію пацієнта. Важливим є також забезпечення зручного інтерфейсу, який дозволить лікарям легко інтерпретувати результати аналізу та приймати рішення на їх основі. Особливу увагу слід приділити питанням безпеки даних, оскільки генетична інформація є дуже чутливою.

На завершення, розробка технології, що використовує SComatic як базовий інструмент, спрямована на вирішення одразу кількох завдань: автоматизації діагностичного процесу, покращення точності прогнозів та інтеграції даних у рамках персоналізованої медицини. Це підкреслює її актуальність у світлі сучасних викликів медичної інформатики та геноміки.

У сучасній медицині, де інформатика й біотехнології стрімко розвиваються, надзвичайно важливим є розробка платформ, які дозволяють забезпечувати масштабовану й високоточну підтримку діагностичних рішень. Однією з таких платформ є підхід до обробки та аналізу мутацій, представлений у статті Nature Biotechnology. Він акцентує увагу на тому, що інтеграція багатопрофільних даних, таких як дані про соматичні мутації, є основою для успішного впровадження персоналізованої медицини. Зокрема, використання методів секвенування нового покоління (NGS) забезпечує величезну кількість даних про геном пацієнта, які, проте, потребують складної обробки для їх коректного використання у клінічній практиці.

У статті також підкреслюється важливість забезпечення високої точності та чутливості алгоритмів для аналізу мутацій. Для цього розробники використовують не лише традиційні методи біоінформатики, але й штучний інтелект та машинне навчання, які дають змогу покращувати інтерпретацію складних даних і зменшувати кількість помилкових результатів. Це особливо

важливо у випадках, коли аналіз мутацій використовується для визначення ризику розвитку хвороб або вибору терапевтичних стратегій. У цьому контексті можливості SComatic для детекції соматичних мутацій та їхнього впливу на функції білків набувають ще більшої актуальності.

Додатково, з огляду на інформацію з джерела, сучасні системи повинні бути адаптованими до роботи в умовах значного обсягу даних, де забезпечення швидкості обробки стає критичним фактором. Інструменти, які можуть забезпечувати автоматизацію цього процесу, мають бути здатними до інтеграції з існуючими електронними медичними записами та іншими платформами для управління клінічними даними. Це дозволяє не лише покращити діагностичний процес, але й інтегрувати геномні знання у реальний клінічний контекст, підвищуючи якість медичних послуг.

Стаття також наголошує на значенні стандартизації даних та забезпеченні прозорості алгоритмів, що використовуються у медичних інформаційних системах. Це дозволяє підвищити довіру до автоматизованих рішень, особливо серед клініцистів, які часто покладаються на такі системи у прийнятті важливих рішень. SComatic, як приклад інструмента з чітко визначеними алгоритмами аналізу мутацій, може слугувати базою для побудови комплексного фреймворку, що включатиме не лише інструмент діагностики на основі мутацій (Mutation-Based Diagnostic Tool), але й модулі для оцінки прогнозів ризиків (Prognostic Risk Assessment), планування терапії за допомогою CRISPR (CRISPR-Driven Precision Therapy Planner) та інтеграції з електронними медичними записами (Integration into an Electronic Health Record).

Таким чином, розробка таких технологій, як SComatic, стає основою для нового етапу розвитку медичних інформаційних систем, де автоматизація, точність та інтеграція даних дозволяють зробити медицину більш персоналізованою, доступною та ефективною.

### 3.2 Архітектура фреймворку на основі SComatic

Архітектура фреймворку на основі SComatic формується як багаторівнева система, що інтегрує інструменти для аналізу геномних даних та підтримки клінічних рішень. Вона побудована з урахуванням потреб сучасних медичних інформаційних систем, включаючи обробку великої кількості даних, модульність, масштабованість та здатність до інтеграції з існуючими електронними медичними записами. Основою для цієї архітектури є SComatic — інструмент, який забезпечує високоточний аналіз соматичних мутацій, використовуючи сучасні алгоритми та методи обчислень, і виступає базою для розробки нових модулів, які розширюють його функціонал.

Першим і ключовим компонентом фреймворку є інструмент діагностики на основі мутацій. Цей модуль забезпечує обробку вхідних даних, які включають результати секвенування геному, клінічні характеристики пацієнта та інші біомедичні дані. Головною метою інструменту є виявлення ключових мутацій, що мають вплив на розвиток патологій, а також надання чіткої та структурованої інформації для клініцистів. Завдяки інтеграції алгоритмів машинного навчання, модуль здатний розпізнавати патерни, які вказують на високий ризик захворювання, навіть у випадках, коли мутації не є очевидно пов'язаними з певною хворобою.

Архітектура діагностичного інструменту передбачає кілька функціональних рівнів. Перший рівень відповідає за попередню обробку даних, де відбувається фільтрація низькоякісних зчитувань та нормалізація інформації. Далі працює аналітичний рівень, який використовує алгоритми для ідентифікації значущих мутацій. На цьому етапі застосовується база знань про мутації, яка включає клінічно перевірену інформацію з таких ресурсів, як ClinVar, COSMIC або HGMD. Останнім рівнем є рівень інтерпретації, який формує рекомендації для лікаря, вказуючи можливі діагностичні сценарії або необхідність додаткових

досліджень.

Інструмент діагностики також інтегрує дані про мутації з клінічними показниками, створюючи звіти, які враховують мультидисциплінарний підхід до діагностики. Наприклад, при виявленні мутацій, пов'язаних із онкогенними шляхами, система може порівняти ці дані з демографічною інформацією, анамнезом пацієнта та попередніми результатами досліджень, щоб запропонувати максимально індивідуалізоване діагностичне рішення.

Крім того, інструмент має можливість навчання та адаптації на основі нових даних, що забезпечує його постійну актуальність у швидкоплинному світі біомедицини. Інтеграція з іншими модулями фреймворку, такими як планувальник CRISPR-редагування або оцінка ризиків, підсилює його функціональність і дозволяє використовувати результати діагностики для планування терапії та довгострокового моніторингу пацієнтів.

Таким чином, інструмент діагностики на основі мутацій виступає центральним елементом фреймворку, забезпечуючи як високу точність аналізу, так і ефективну комунікацію результатів з іншими модулями та кінцевими користувачами, що робить його невід'ємною частиною сучасної персоналізованої медицини.

Прогностична оцінка ризиків є одним із ключових компонентів розроблюваного фреймворку, що базується на SComatic, і відіграє важливу роль у забезпеченні персоналізованого підходу до лікування пацієнтів. Цей модуль спрямований на аналіз даних про генетичні мутації з урахуванням клінічних та демографічних показників для прогнозування ймовірності розвитку захворювань, їх прогресування та можливих ускладнень. Прогностична оцінка допомагає лікарям приймати обґрунтовані рішення щодо подальшого спостереження пацієнта або застосування профілактичних заходів.

Головна особливість модуля полягає в інтеграції даних про мутації з результатами епідеміологічних досліджень, біомедичних баз даних та алгоритмами машинного навчання. Наприклад, виявлення мутацій у генах,

пов'язаних із підвищеним ризиком онкологічних захворювань (BRCA1/BRCA2 або TP53), дозволяє створювати персоналізовані прогностичні моделі для кожного пацієнта. Такі моделі можуть враховувати генетичну схильність, зовнішні фактори, включаючи спосіб життя, і специфічні біомаркери, що забезпечує високу точність прогнозів.

Цей модуль працює на основі багаторівневого підходу. На першому етапі проводиться класифікація мутацій за ступенем їхньої клінічної значущості з використанням баз даних, таких як ClinVar, COSMIC або HGMD. Другий етап включає аналіз зв'язків між виявленими мутаціями і прогнозованими клінічними результатами за допомогою машинного навчання та алгоритмів глибокого навчання. Зокрема, використовуються моделі, які можуть навчатися на великих наборах даних і враховувати складні взаємодії між множинними змінними.

Прогностична оцінка ризиків також дозволяє визначати групи пацієнтів із найвищими ризиками розвитку певних станів, що є критично важливим для раннього втручання. Наприклад, пацієнти з певними мутаціями можуть бути спрямовані на більш часті обстеження або запропоновані їм профілактичні заходи, такі як медикаментозна терапія чи зміни способу життя. Крім того, прогностичні моделі можуть бути використані для оцінки ефективності різних терапевтичних стратегій, що допомагає лікарям обирати найоптимальніший план лікування.

Інтеграція цього модуля в електронні медичні записи дозволяє автоматично оновлювати прогностичну інформацію у міру появи нових даних, зберігаючи актуальність прогнозів. У поєднанні з іншими модулями фреймворку, такими як діагностичний інструмент чи планувальник терапії, прогностична оцінка ризиків забезпечує комплексний підхід до медичного обслуговування, підвищуючи ефективність діагностики та лікування.

Отже, розробка модуля прогностичної оцінки ризиків не лише допомагає у виявленні пацієнтів із високим ризиком захворювань, але й формує основу для прийняття клінічних рішень, що сприяє переходу до персоналізованої медицини.

Планувальник прецизійної терапії з використанням CRISPR (CRISPR-Driven Precision Therapy Planner) є одним із найбільш інноваційних компонентів фреймворку, орієнтованого на підтримку діагностичних рішень у медичних інформаційних системах. Його розробка передбачає створення інтерактивного інструменту, здатного автоматизувати аналіз та оптимізацію терапевтичних стратегій, заснованих на редагуванні генома з використанням технології CRISPR.

Основна задача планувальника полягає у визначенні найоптимальніших шляхів редагування геному, спрямованих на корекцію мутацій, які є причиною розвитку патологій. У цьому контексті CRISPR виступає як інструмент молекулярного рівня, що дозволяє точно модифікувати послідовності ДНК. З точки зору комп'ютерної науки, функціонування планувальника базується на реалізації алгоритмів оптимізації, моделювання та прогнозування, які працюють із великими обсягами даних генетичних послідовностей.

Архітектурні компоненти планувальника:

1) база даних геномних послідовностей і мутацій. У процесі роботи планувальника використовується інтегрована база даних, що містить геномні послідовності пацієнта та відомості про виявлені мутації. Дані зберігаються у вигляді індексованих структур, таких як бази даних BLAST чи аналогічні сховища, що забезпечують швидкий доступ до інформації. Це дозволяє проводити аналіз мутацій у режимі реального часу;

2) модуль аналізу цільових ділянок CRISPR. Основою цього компонента є алгоритми пошуку та вибору CRISPR-мішеней. Такі алгоритми, як CHOPCHOP або CRISPR-Cas9 Target Finder, виконують аналіз геномних послідовностей для визначення найбільш ефективних ділянок для редагування. У цьому процесі враховуються такі критерії, як ефективність зв'язування гідів (guide RNAs) та мінімізація позацільових ефектів;

3) інтеграція з електронними медичними записами (EMR). Планувальник

отримує клінічні дані пацієнтів безпосередньо з EMR, включаючи історію хвороби, біохімічні показники та результати попередніх терапій. Ці дані автоматично інтегруються в модель, що дозволяє адаптувати рекомендації під конкретного пацієнта.

Наш реалізований планувальник для аналізу мутацій базується на інтерактивному інтерфейсі, створеному за допомогою Streamlit, який забезпечує зручність використання та швидкий доступ до результатів аналізу. Система дозволяє завантажувати мутаційні дані, виконувати оцінку ризиків, визначати потенційні мішені CRISPR та створювати звіти у форматах PDF або HTML. Інтерфейс інтегрує візуалізацію результатів, використовуючи гістограми, графіки розподілу ризиків і таблиці, що полегшує розуміння отриманих результатів навіть для нефахівців.

У планувальнику також реалізовано обчислення ризику на основі мутаційних даних. Алгоритм оцінює ризик залежно від функцій мутацій та їхньої локалізації, а також генів, які беруть участь у змінах. Це дозволяє визначати мутації, що мають високий клінічний ризик, та забезпечувати точні рекомендації для подальшого прийняття рішень. Виявлення мішеней CRISPR виконується шляхом аналізу послідовностей ДНК із застосуванням правил, що враховують специфічність і ефективність генного редагування.

Попри досягнуті результати, реалізована система не використовує розподілені обчислення, такі як Apache Hadoop або Spark, для обробки великих обсягів даних. Це можна впровадити в майбутньому для аналізу великих геномних баз даних. Крім того, індексація послідовностей і обчислення на GPU, які могли б значно прискорити пошук мішеней CRISPR, поки що не інтегровані, але залишаються перспективним напрямом для вдосконалення.

Хмарні сервіси, такі як AWS чи Google Cloud, також не застосовувалися на даному етапі, проте їх використання в майбутньому забезпечить масштабованість і гнучкість системи, дозволяючи адаптувати обчислювальні ресурси до обсягів даних. Наразі обробка даних здійснюється локально, що є

достатнім для тестового використання, але може стати обмеженням при роботі з великими обсягами інформації.

Безпека даних пацієнтів поки не була реалізована, але в майбутньому планується використання методів шифрування, таких як AES-256, а також диференціальної приватності для аналізу даних без розкриття індивідуальних характеристик. Ці заходи є критично важливими для інтеграції системи в клінічну практику.

Планувальник об'єднує підходи з галузей біоінформатики, обчислювальної біології, машинного навчання та медичних інформаційних систем. Це сприяє розвитку мультидисциплінарних досліджень і стимулює появу нових рішень для прецизійної медицини.

Інтеграція планувальника в загальний фреймворк підтримки діагностичних рішень відкриває перспективи для розробки нових модулів, таких як інструменти прогнозування довгострокових ефектів терапії чи моделі для оптимізації лікування на популяційному рівні. Це може сприяти створенню повноцінних екосистем для персоналізованої медицини.

Інтеграція інструментів підтримки діагностичних рішень, таких як планувальник прецизійної терапії на основі CRISPR, в електронні медичні карти (ЕМК) є надзвичайно важливим етапом у розвитку медичних інформаційних технологій. Ця інтеграція дозволяє об'єднати клінічні, геномні та інші дані про пацієнта в єдину систему, що забезпечує лікарям можливість отримувати комплексні та актуальні дані для прийняття рішень щодо лікування.

Першим важливим аспектом інтеграції є здатність синхронізувати дані між різними джерелами, такими як лабораторні аналізи, генетичні дослідження, попередні медичні звіти та поточні клінічні записи пацієнта. Використання стандартів, таких як HL7 або FHIR (Fast Healthcare Interoperability Resources), є основою для забезпечення сумісності між різними системами та даними. Ці стандарти дозволяють системам ЕМК обмінюватися інформацією без втрат та помилок, що є критично важливим для надійності та безпеки даних.

Інтеграція планувальника CRISPR у структуру ЕМК дозволяє не тільки враховувати наявні генетичні мутації пацієнта, а й обчислювати прогнозовані ефекти терапевтичного втручання. Збір та обробка даних в режимі реального часу забезпечує своєчасний доступ до необхідної інформації. Наприклад, після виконання генетичного тесту система автоматично завантажує результати у карту пацієнта, де вони відразу можуть бути проаналізовані за допомогою відповідних інструментів для планування терапії. Таке рішення дозволяє лікарям отримувати рекомендації по лікуванню, засновані на індивідуальних характеристиках кожного пацієнта.

Особливо важливим є забезпечення безпеки та конфіденційності даних. Враховуючи високий рівень чутливості генетичної інформації, система повинна використовувати сучасні методи шифрування та захисту даних, що гарантують збереження персональної інформації пацієнтів. Для цього застосовуються як традиційні методи шифрування даних, так і більш складні технології, такі як блокчейн, які дозволяють створювати незмінні та прозорі записи в медичній картці.

Інтеграція також має значний вплив на автоматизацію процесу прийняття рішень. Лікарям надається більше часу для консультування та лікування пацієнтів, оскільки система автоматично обробляє великі обсяги даних та надає рекомендації щодо терапії. Це не лише полегшує роботу лікаря, а й підвищує якість медичних послуг, зменшуючи ймовірність помилок, що можуть виникнути внаслідок людського фактору.

У перспективі інтеграція таких інструментів у ЕМК відкриває нові можливості для розвитку прецизійної медицини. Вона дозволяє зберігати всю історію лікування пацієнта, включаючи генетичні аналізи, що надасть можливість для довгострокового моніторингу ефективності лікування та адаптації терапії в залежності від змін у геномі. Такий підхід є важливим для розвитку індивідуалізованого лікування, де кожен пацієнт отримує найефективніше для нього лікування на основі його генетичних особливостей та

медичної історії.

Завдяки інтеграції в систему ЕМК, можливе також автоматичне оновлення терапевтичних рекомендацій на основі нових наукових даних або змін у клінічному стані пацієнта. Це дозволяє лікарям постійно підтримувати актуальність інформації та забезпечує кращу реакцію на зміни у здоров'ї пацієнта. Крім того, використання даних з ЕМК допомагає створювати більш точні моделі для прогнозування результатів лікування та ризиків розвитку захворювань у майбутньому.

Таким чином, інтеграція технології підтримки діагностичних рішень на основі CRISPR в електронні медичні карти є потужним кроком у напрямку розвитку персоналізованої медицини, яка враховує як генетичні фактори, так і клінічні характеристики пацієнта для прийняття найефективніших рішень щодо лікування.

### 3.3 Технологічні аспекти реалізації

Технологічні аспекти реалізації фреймворку для підтримки діагностичних рішень на основі мутаційного аналізу є складним і багатограним процесом, що включає в себе вибір відповідних алгоритмів, мов програмування, баз даних, а також стратегій інтеграції з існуючими медичними інформаційними системами. Важливим етапом є те, що інструмент SComatic, завдяки своїй гнучкості, можна адаптувати до різних задач, зокрема до вимог конкретних медичних установ, що дає змогу розширювати його функціональність за потреби. Ця адаптивність є особливо цінною в умовах, коли необхідно забезпечити інтероперабельність між різними програмними рішеннями та інтеграцію з іншими інструментами для аналізу даних.

Один із ключових моментів, на який слід звернути увагу, — це

використання алгоритмів для аналізу мутацій та прогнозування терапевтичних рішень. Для цього SComatic може застосовувати різноманітні методи обробки даних, такі як машинне навчання та статистичний аналіз, що дозволяє не тільки виявляти мутації, але й оцінювати їх вплив на розвиток хвороб та відповіді на терапію. Алгоритми можуть варіюватися від простих методів фільтрації та сортування до складних моделей, заснованих на глибокому навчанні, що дають змогу робити більш точні прогнози для кожного пацієнта. Мови програмування, які використовуються для реалізації цих алгоритмів, зазвичай включають Python, R та C++, оскільки вони підтримують широкий спектр бібліотек для аналізу даних і дозволяють ефективно працювати з великими обсягами інформації, необхідними для геномного аналізу.

Бази даних, що використовуються для зберігання та обробки генетичних даних, повинні бути здатні ефективно працювати з великими наборами інформації та забезпечувати високу швидкість доступу до даних, а також надійність і безпеку збереження персональних даних пацієнтів. Часто для таких цілей використовуються реляційні бази даних, такі як MySQL або PostgreSQL, або NoSQL рішення, такі як MongoDB, коли потрібна більша гнучкість у зберіганні та доступі до даних. Крім того, використання хмарних технологій дозволяє забезпечити масштабованість та доступність даних, що є важливим у контексті медичних досліджень і надання послуг для великої кількості пацієнтів.

Важливою складовою реалізації є інтеграція SComatic з існуючими медичними системами, зокрема з електронними медичними картами (ЕМК) та іншими інформаційними системами охорони здоров'я. Це дозволяє лікарям і медичним працівникам отримувати повну картину стану пацієнта, враховуючи як клінічні, так і генетичні дані. Для інтеграції застосовуються відкриті стандарти, такі як FHIR (Fast Healthcare Interoperability Resources) та HL7, які забезпечують можливість обміну даними між різними системами без втрат інформації. Це дає змогу автоматично оновлювати медичні карти пацієнтів на основі нових результатів аналізів або змін у стані здоров'я, знижуючи ризик

помилки та забезпечуючи лікарів необхідною інформацією в реальному часі.

Важливим аспектом є також забезпечення безпеки та конфіденційності даних. Генетична інформація є дуже чутливою, і її захист має критичне значення. Для цього використовуються сучасні методи шифрування та аутентифікації користувачів, а також механізми для контролю доступу, щоб тільки авторизовані особи могли отримувати доступ до персональних даних пацієнтів. Враховуючи величезний обсяг генетичних та клінічних даних, необхідно забезпечити їх ефективне зберігання та швидкий доступ до них, що реалізується через використання розподілених обчислювальних систем та технологій обробки великих даних (Big Data).

Крім того, важливою складовою є інтеграція нових наукових досягнень в області геноміки та терапії, таких як CRISPR. Використання інструментів для аналізу даних та планування терапії на основі CRISPR вимагає постійного оновлення алгоритмів та баз даних з новими науковими відкриттями, що дозволяє підвищити точність прогнозів і ефективність лікування. Це також створює додаткові вимоги до постійного моніторингу та адаптації системи до нових даних.

Таким чином, технологічні аспекти реалізації фреймворку для підтримки діагностичних рішень на основі мутаційного аналізу є надзвичайно важливими для забезпечення ефективності та надійності цієї системи. Вибір правильних алгоритмів, мов програмування, баз даних і методів інтеграції з іншими системами є критичним для досягнення високої точності діагностики та надання пацієнтам персоналізованої допомоги на основі їх генетичних характеристик.

Однією з основних переваг запропонованого фреймворку на основі SComatic є його здатність автоматизувати процеси діагностики та прогнозування на основі мутаційного аналізу, що робить його значно ефективнішим і швидшим порівняно з традиційними методами. Сучасні медичні системи зазвичай мають складнощі в інтеграції різноманітних даних з різних джерел, таких як геномні, клінічні та демографічні дані, що значно ускладнює прийняття рішень в умовах

персоналізованої медицини. Однак фреймворк, розроблений на основі SComatic, дозволяє вирішити це питання, об'єднуючи всі необхідні дані в єдину систему та автоматизуючи процес їх аналізу.

Інтеграція даних з різних джерел у рамках єдиної платформи є однією з головних переваг цієї розробки. Більшість існуючих медичних інформаційних систем працюють із застарілими протоколами і часто не здатні безперешкодно обробляти великі обсяги даних, що виходять із сучасних геномних досліджень. Проте розроблений фреймворк включає можливість гнучкого оброблення і інтеграції цих даних, використовуючи сучасні підходи до управління великими даними та інтероперабельності систем, що дозволяє системі працювати більш ефективно.

Важливим аспектом є орієнтованість фреймворку на персоналізовану медицину. Завдяки використанню технологій, таких як CRISPR і аналіз мутацій, система може створювати індивідуальні рекомендації щодо лікування для кожного пацієнта, враховуючи його унікальний генетичний профіль. Це дозволяє не лише значно покращити точність діагностики, але й оптимізувати лікування, підвищуючи його ефективність і знижуючи ризики ускладнень.

Гнучкість системи є ще однією ключовою перевагою, оскільки вона дозволяє адаптувати її до потреб різних медичних установ і різних типів пацієнтів. Враховуючи швидкий розвиток геномних технологій, фреймворк забезпечує простоту оновлення та інтеграції нових алгоритмів і даних, що дає змогу вчасно впроваджувати останні досягнення в науці. У той же час, фреймворк забезпечує високу сумісність з іншими медичними системами, що дозволяє інтегрувати його в існуючі інфраструктури без великих змін у вже налаштованих процесах.

При порівнянні з іншими існуючими розробками в галузі медичної діагностики, SComatic та його адаптовані версії в рамках цього фреймворку виділяються своєю універсальністю та потужністю. Вони не лише автоматизують процеси обробки даних, а й забезпечують складну інтерпретацію

генетичних мутацій, їх взаємодій і потенційного впливу на хвороби, що є важливим кроком уперед у галузі персоналізованої медицини.

Щоб порівняти запропонований фреймворк на основі SComatic з іншими існуючими розробками, варто розглянути кілька відомих інструментів, які вже активно використовуються в медичних системах для діагностики на основі мутацій. Більшість з них фокусуються на певних аспектах генетичної діагностики та не завжди враховують комплексний підхід, який включає інтеграцію з клінічними даними, прогнозування ризиків і планування терапії, що є сильними сторонами розробленого фреймворку.

Одним із таких інструментів є OncoKB — база даних, яка класифікує генетичні мутації і їх значення для раку. Вона допомагає лікарям виявляти мутації, що впливають на лікування пацієнтів, але ця система обмежується лише класифікацією мутацій і не має розвинених інструментів для інтеграції з іншими медичними даними або для розробки індивідуальних терапевтичних планів. OncoKB орієнтується на статичні дані і не дозволяє автоматизувати процеси на основі поточних медичних умов кожного конкретного пацієнта. Крім того, вона не підтримує інтеграцію з електронними медичними картами, що значно ускладнює її використання в реальних умовах клінічної практики.

Інший приклад — це систему GATK (Genome Analysis Toolkit), яка надає потужні інструменти для виявлення генетичних варіацій. Хоча GATK широко використовується для аналізу великих геномних даних, її функціональність обмежена лише на етапі варіантного аналізу, і вона не інтегрується безпосередньо з клінічними або терапевтичними даними пацієнтів. Більш того, GATK не пропонує інструменти для прогнозування ризику розвитку захворювань або для розробки планів лікування на основі генетичних мутацій, що є значною перевагою запропонованої розробки.

Програма Cancer Genome Interpreter — це ще один інструмент, що дозволяє ідентифікувати мутації, які можуть вплинути на діагноз та терапевтичні рішення для пацієнтів з раком. Проте, подібно до OncoKB, ця система не забезпечує

інтеграцію з іншими медичними даними, такими як клінічна історія пацієнта, і не пропонує можливості для прогнозування ризиків чи планування терапії на основі індивідуальних мутацій. Таким чином, хоча вона добре працює в рамках мутаційного аналізу, вона не може в повній мірі замінити комплексні платформи, орієнтовані на персоналізовану медицину.

У таблиці 3.1 показано порівняння базового методу з іншими.

Таблиця 3.1 – Порівняння SComatic із іншими інструментами

Критерій	Базовий метод SComatic	OncoKB	GATK	Cancer Genome Interpreter
Автоматизація процесів	Повна автоматизація	Лише класифікація	Лише варіантний аналіз	Лише діагностичний аналіз
Інтеграція з клінічними даними	Підтримується	Немає	Немає	Немає
Індивідуальне лікування	Прогнозування і терапія	Не підтримується	Не підтримується	Не підтримується
Гнучкість	Висока (оновлення алгоритмів)	Обмежена	Обмежена	Обмежена
Швидкість роботи	Оптимізована	Висока	Висока	Висока

На відміну від цих інструментів, фреймворк на основі SComatic забезпечує не тільки точну і швидку ідентифікацію мутацій, а й інтеграцію з клінічними даними, що дозволяє створювати персоналізовані стратегії лікування. Крім того,

він включає можливості для прогнозування ризиків та автоматизованого планування терапії на основі CRISPR, що робить його більш гнучким і ефективним у порівнянні з іншими розробками [26, 28], як показано у таблиці 3.1.

З інтеграцією в електронну медичну картку система дозволяє лікарям приймати обґрунтовані рішення в режимі реального часу, що істотно покращує якість медичних послуг і знижує ймовірність помилок.

Таким чином, фреймворк SComatic є значним кроком уперед у порівнянні з існуючими аналогами, які обмежені тільки варіантним аналізом або не мають належної інтеграції з іншими медичними даними.

## 4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА ОТРИМАНИХ НАУКОВИХ РЕЗУЛЬТАТІВ

### 4.1 Розробка модулю підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу

Розробка модуля для аналізу мутацій у рамках підтримки діагностичних рішень потребує структурованого підходу до роботи з даними, особливо коли йдеться про обробку генетичної інформації. У першому етапі цього процесу основний акцент був зроблений на завантаженні, інтеграції та попередній обробці даних із використанням інструменту SComatic. Ця програма забезпечує генерацію високоякісних вихідних файлів, таких як *BaseCellCountsMerged* та *VariantCalling*, які містять інформацію про соматичні мутації та базові підрахунки для кожної клітини, що аналізується. Робота з такими даними вимагає чіткого розуміння їхньої структури та логіки, адже саме на цьому етапі формується основа для подальшого аналізу.

Об'єднання даних, зокрема злиття інформації про мутації із відповідними базовими підрахунками, виконувалося за спільними ідентифікаторами, такими як хромосомні координати (*CHROM* та *Start*). Це дозволило поєднати різноманітні джерела даних у єдину таблицю, яка забезпечує цілісний погляд на кожну мутацію в контексті її геномного оточення (рис. 4.1).

Ця процедура була автоматизована з використанням мови Python та бібліотек для обробки даних, таких як Pandas. Такий підхід не лише мінімізує ризик помилок, але й забезпечує можливість масштабування аналізу на великі обсяги даних, що є типовими для сучасних генетичних досліджень.

Одним із ключових аспектів цього етапу стало створення нової метрики під назвою *Impact Score*, яка дозволяє кількісно оцінити потенційний вплив мутацій. Формула для обчислення цього показника (4.1) враховує частоту варіантного алеля (*VAF*) та суму базових підрахунків для аденіну, тиміну,

цитозину та гуаніну в стромальних клітинах. Наприклад, якщо частота варіантного алеля становить 0.25, а базові підрахунки для стромальних клітин дорівнюють 100, *Impact Score* становитиме 25. Цей показник є важливим інструментом для автоматизації процесу пріоритетизації мутацій, оскільки дозволяє швидко ідентифікувати мутації з найбільшим потенційним впливом.

```

  CHROM      Start      End_x REF_x ALT FILTER Cell_types Up_context \
0 chr10  29555105  29555105    G  .      .      .      ATTCT
1 chr10  29555106  29555106    C  .      .      .      TTCTG
2 chr10  29555107  29555107    A  .      .      .      TCTGC
3 chr10  29555108  29555108    A  .      .      .      CTGCA
4 chr10  29559429  29559429    T  .      .      .      TAAAT

Down_context N_ALT  ...          INFO_x Myeloids_x \
0      CAACT  .  ...  DP|NC|CC|BC|BQ|BCf|BCr      NaN
1      AACTG  .  ...  DP|NC|CC|BC|BQ|BCf|BCr      NaN
2      ACTGG  .  ...  DP|NC|CC|BC|BQ|BCf|BCr      NaN
3      CTGGA  .  ...  DP|NC|CC|BC|BQ|BCf|BCr      NaN
4      CTTCA  .  ...  DP|NC|CC|BC|BQ|BCf|BCr      NaN

Epithelial_cells_x          Stromal_cells_x \
0      NaN  5|5|0:0:0:5:0:0|0:0:0:5:0:0|0:0:0:196:0:0|0:0:...
1      NaN  5|5|0:5:0:0:0:0|0:5:0:0:0:0|0:197:0:0:0:0|0:0:...
2      NaN  5|5|5:0:0:0:0:0|5:0:0:0:0:0|192:0:0:0:0:0|0:0:...
3      NaN  7|7|7:0:0:0:0:0|7:0:0:0:0:0|287:0:0:0:0:0|0:0:...
4      NaN  7|6|0:0:6:0:0:0|0:0:7:0:0:0|0:0:274:0:0:0|0:0:...

      End_y REF_y          INFO_y Myeloids_y Epithelial_cells_y \
0  29555105    G  DP|NC|CC|BC|BQ|BCf|BCr      NaN      NaN
1  29555106    C  DP|NC|CC|BC|BQ|BCf|BCr      NaN      NaN
2  29555107    A  DP|NC|CC|BC|BQ|BCf|BCr      NaN      NaN
...
3  7|7|7:0:0:0:0:0|7:0:0:0:0:0|287:0:0:0:0:0|0:0:...
4  7|6|0:0:6:0:0:0|0:0:7:0:0:0|0:0:274:0:0:0|0:0:...

```

Рисунок 4.1 – Об’єднана таблиця з геномними даними (тільки перші 5 строк виведені на екран для демонстраційних цілей).

$$\text{Impact\_Score}_i = \text{VAF}_i \cdot (\text{A\_stromal}_i + \text{T\_stromal}_i + \text{C\_stromal}_i + \text{G\_stromal}_i), \quad (4.1)$$

де  $\text{VAF}_i$  – варіантна алельна частота для  $i$ -го запису;

$\text{A\_stromal}_i$ ,  $\text{T\_stromal}_i$ ,  $\text{C\_stromal}_i$ ,  $\text{G\_stromal}_i$  – значення відповідних

стромальних показників для  $i$ -го запису;

$\text{Impact\_Score}_i$  – результуючий впливовий бал для  $i$ -го запису.

Для забезпечення високої точності аналізу були застосовані методи фільтрації, які дозволили видалити мутації з низькими значеннями *Impact Score*, що не мають істотного біологічного чи клінічного значення. Крім того, дані було відсортовано у спадному порядку за значенням метрики, що забезпечило чітку ієрархію мутацій за ступенем їхнього впливу. Це особливо корисно у випадках, коли дослідник або клініцист працює з великими наборами даних і потребує швидкого доступу до найважливішої інформації.

Для покращення інтерпретації результатів було створено візуалізацію, яка включала графік із 10 найвищими значеннями *Impact Score* (рис. 4.2).

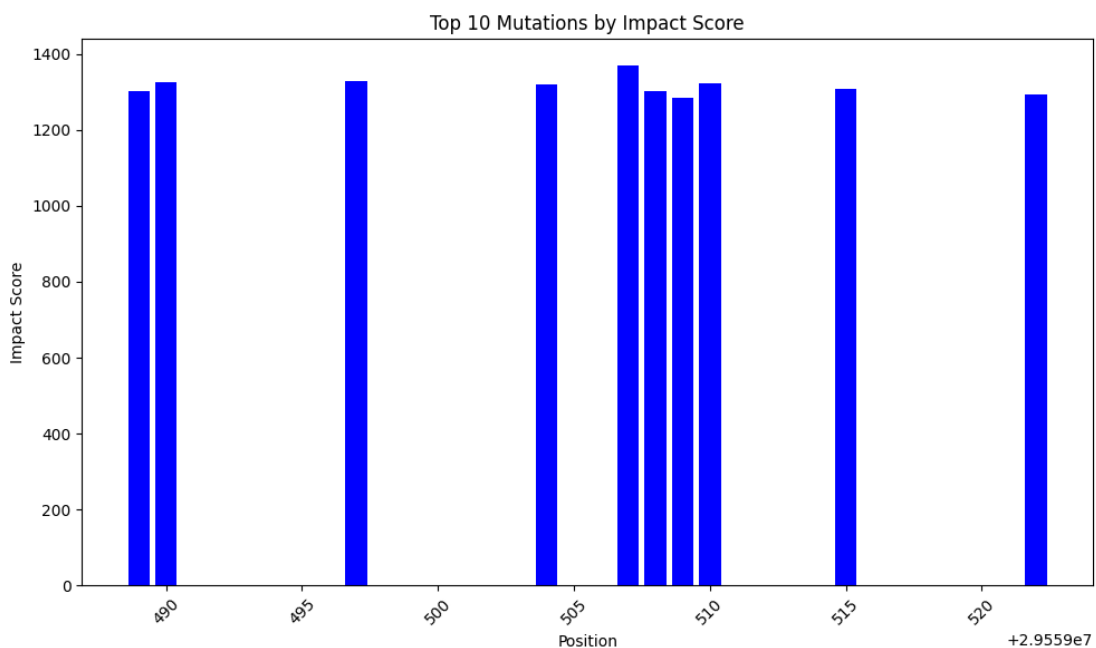


Рисунок 4.2 – Візуалізація топ-10 значень *Impact Score*

Візуалізація даних є важливим інструментом у комп'ютерних науках, адже вона сприяє кращому розумінню результатів аналізу та дозволяє ефективно комунікувати отримані висновки. Наприклад, використовуючи бібліотеки Matplotlib і Seaborn у Python, було створено бар-чарт, що демонструє значення

*Impact Score* для кожної з пріоритетних мутацій. Такий підхід дозволяє не лише побачити, які мутації є найбільш значущими, але й оцінити можливі патерни їхньої присутності в різних геномних регіонах.

Заключним етапом стала генерація вихідного CSV-файлу, який містить усі оброблені дані, включаючи розраховані значення *Impact Score*. Цей файл є основою для подальших етапів аналізу, таких як оцінка ризиків, створення прогнозних моделей або розробка систем підтримки рішень. Використання універсального формату CSV гарантує сумісність із різними інструментами для аналізу даних, включаючи такі програми, як RStudio або SAP Analytics Cloud.

Ця частина роботи є критично важливою з точки зору майбутніх розрахунків, оскільки забезпечує підготовку даних до складніших етапів аналізу. Чисті, структуровані та пріоритизовані дані є ключем до ефективної розробки алгоритмів, побудови моделей машинного навчання та прийняття рішень у режимі реального часу. Таким чином, цей крок не лише формує технічну базу для подальших досліджень, але й закладає основи для інтеграції отриманих результатів у клінічну практику.

Подивимось на реалізацію наступного етапу мутаційного аналізу після виконаного препроцесингу.

Оцінка прогностичного ризику у контексті аналізу мутацій є одним із ключових етапів у створенні систем підтримки діагностичних рішень, оскільки вона дозволяє не лише ідентифікувати мутації, але й класифікувати їх за ступенем потенційного впливу на клінічний результат. Цей процес базується на розробці ризикової моделі, яка використовує конкретні критерії для оцінки кожної мутації, забезпечуючи їхній поділ на категорії ризику: «Без ризику», «Низький ризик», «Помірний ризик» та «Високий ризик». Використання формального підходу до розробки моделі дозволяє автоматизувати цей процес, підвищуючи його точність і масштабованість.

Модель оцінки ризику базується на наборі правил, які враховують критичність генів, функціональні наслідки мутацій і їхню локалізацію.

Наприклад, мутації у відомих генах-супресорах пухлин, таких як *TP53*, отримують вищі оцінки через їхній значний вплив на розвиток онкологічних захворювань. Якщо функціональне навантаження мутації включає зрушення рамки зчитування (*frameshift*) або стоп-кодон (*stopgain*), такі випадки також додають бали до загального ризикового балу. Аналогічно, сплайсингові мутації, які впливають на регуляцію генів, розглядаються як такі, що підвищують ризик. У кінцевому підсумку, загальний бал дозволяє автоматично визначити категорію ризику для кожної мутації.

Дані для оцінки ризику були підготовлені за допомогою попередньо оброблених файлів, які містять інформацію про хромосомні координати, референсні та альтернативні алелі, функціональні анотації та результати ризикової оцінки виповненої за допомогою Annovar (рис. 4.3).



Рисунок 4.3 – Загальна схема обробки файлів в Annovar

Використовуючи Python, було створено скрипт, який обчислює ризикові бали для кожної мутації та додає новий стовпець із категорією ризику до датасету. Наприклад, у випадку мутації, яка локалізована у гені *BRCA1* і викликає зрушення рамки зчитування, їй буде присвоєно статус «Високий ризик», що відображає її потенційно серйозний вплив.

Для забезпечення зручного аналізу результатів був згенерований звіт у вигляді текстового файлу. У цьому звіті представлено ключові характеристики кожної мутації, включаючи їхній функціональний вплив, анотації змін

амінокислот і оцінку ризику. Наприклад, рядок зі звіту може виглядати так: «chrX, 274518, 274521, A, T, exonic, BRCA1, frameshift, p.Val600Ala, High Risk» (рис. 4.4).

Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	GeneDetail.refGene	ExonicFunc.refGene
chrX	430200	430203	A	T	intronic	KRAS	nonsynonymous p.Val205Leu	No Risk
chr13	345499	345500	T	C	splicing	KRAS	nonsynonymous p.Val399Leu	Low Risk
chr2	162617	162619	T	G	splicing	KRAS	p.Val521Ala	Low Risk
chr11	315491	315499	C	A	splicing	BRCA2	stopgain p.Val525Ala	Moderate Risk
chrX	462141	462147	A	G	splicing	TP53	stopgain p.Val288Glu	High Risk
chr2	270490	270496	C	T	intronic		frameshift p.Val438Leu	Low Risk
chr21	474038	474048	G	T	intronic	EGFR	stopgain p.Val389Glu	Low Risk
chr2	137857	137866	C	G	exonic		frameshift p.Val475Ala	Low Risk
chr2	128101	128102	A	G	intronic	BRCA2	p.Val391Leu	No Risk
chr2	230268	230278	G	T	intronic	BRCA2	nonsynonymous p.Val575Glu	No Risk
chr18	446750	446757	A	T	exonic	EGFR	frameshift p.Val362Leu	Low Risk
chr12	335898	335903	G	T	splicing	BRCA1	p.Val97Ala	Low Risk
chr10	389979	389982	T	T	exonic	BRCA2	p.Val451Leu	No Risk
chr27	283169	283173	T	C	intronic	BRCA2	frameshift p.Val119Leu	Low Risk
chr17	235608	235612	T	A	exonic		nonsynonymous p.Val276Glu	No Risk
chr18	234365	234366	A	T	intronic	KRAS	nonsynonymous p.Val345Ala	No Risk
chr11	337379	337385	A	A	exonic	BRCA2	stopgain p.Val398Ala	Low Risk
chr27	498691	498694	C	A	intronic	KRAS	p.Val439Ala	No Risk
chr3	193920	193922	G	T	exonic	BRCA2	frameshift p.Val218Glu	Low Risk

Рисунок 4.4 – Вихідні дані на даному етапі роботи

Така структура забезпечує компактне представлення даних, що полегшує їхню інтерпретацію як дослідниками, так і клініцистами.

Щоб зробити результати більш інтуїтивно зрозумілими, були створені візуалізації, які демонструють розподіл мутацій за категоріями ризику (рис. 4.5).

Використовуючи бібліотеки Matplotlib і Seaborn, було побудовано гістограму, що показує кількість мутацій у кожній категорії, а також кругову діаграму, яка ілюструє частки мутацій різних рівнів ризику. Наприклад, діаграма може показати, що 50% мутацій класифікуються як «Без ризику», тоді як 10% мають статус «Високий ризик». Це дозволяє швидко оцінити загальну картину і визначити, які мутації потребують першочергової уваги (рис. 4.6 та рис. 4.7).

З точки зору комп'ютерних наук, цей етап є важливим не лише через його прикладну значущість, але й через використання передових методів обробки даних і автоматизації процесів.

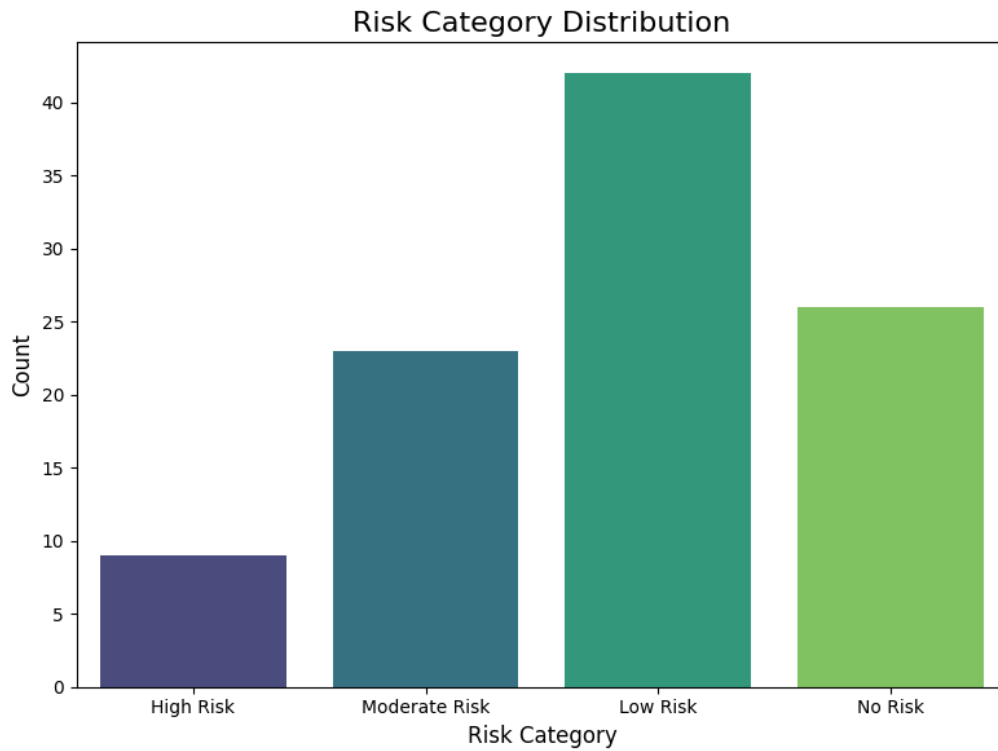


Рисунок 4.5 – Візуалізація частоти мутацій у кожній категорії

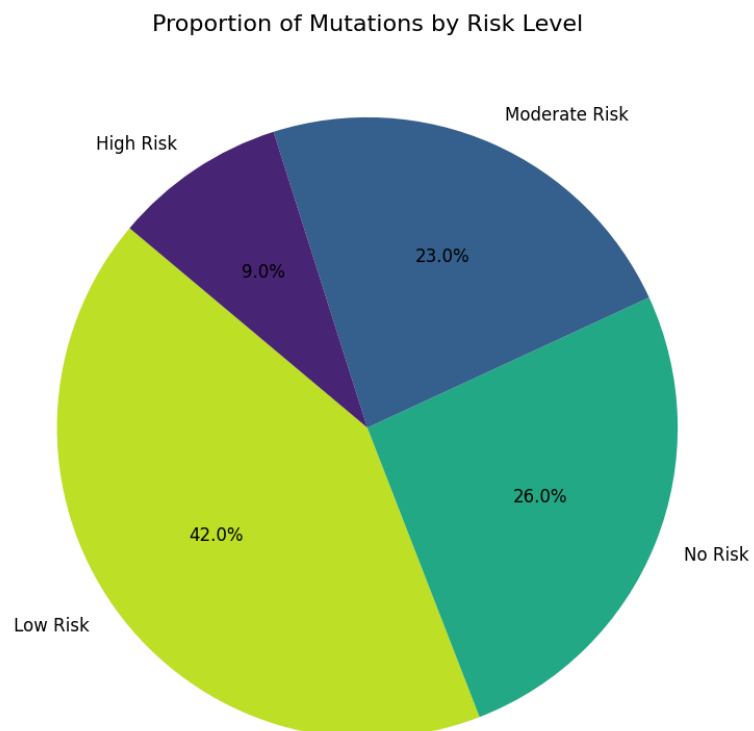


Рисунок 4.6 – Кругова діаграма для першочергової оцінки загальної картини розподілу наявних мутацій за рівнем потенційного ризику

Використання мови Python забезпечує високу продуктивність і гнучкість у роботі з великими наборами даних. Крім того, модульність створеного коду дозволяє легко додавати нові критерії оцінки або інтегрувати зовнішні джерела даних для підвищення точності моделі. У майбутньому така система може бути інтегрована в більш складні алгоритми машинного навчання, що дозволить не лише класифікувати мутації, але й прогнозувати їхній вплив на клінічні результати.

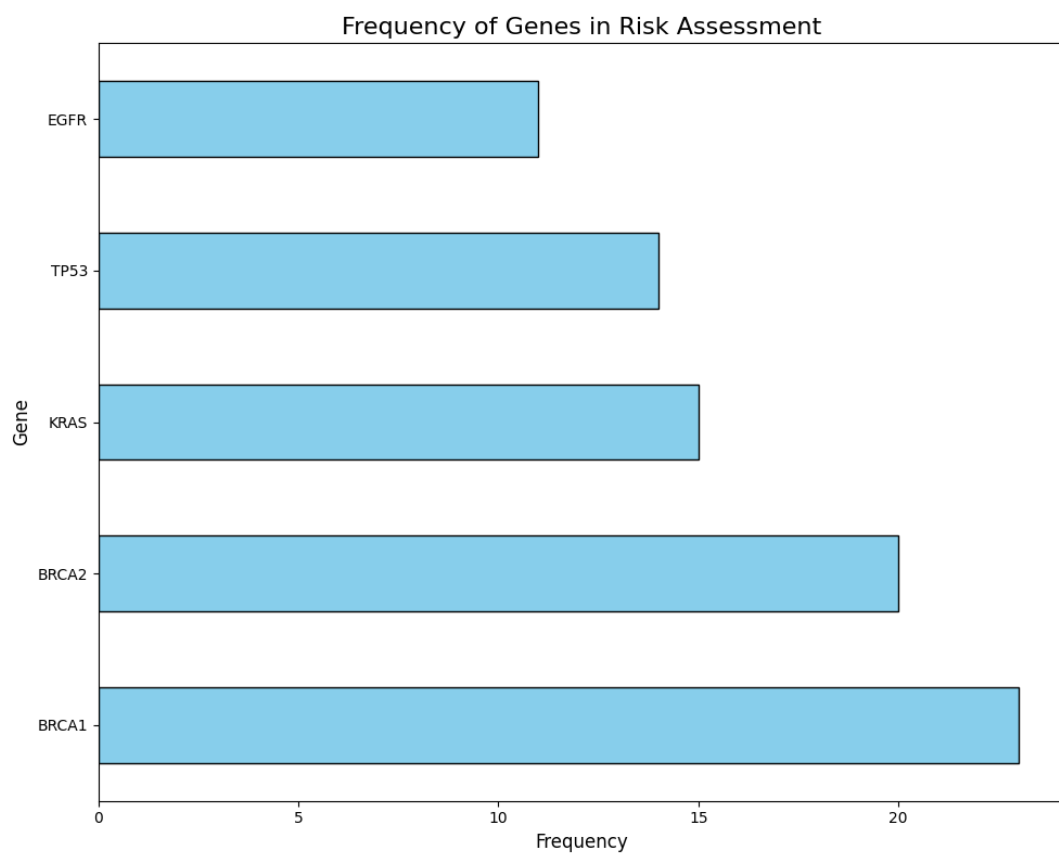


Рисунок 4.7 – Діаграма для оцінки частоти зустрічання мутацій генів, представляючих ризик для пацієнта.

Перейдемо до реалізації третього етапу.

Даний модуль — це приклад інтеграції біоінформатики та комп'ютерних наук для розробки інструментів, що допомагають ідентифікувати та оцінювати мішені для гРНК у терапевтичних цілях. Цей модуль фокусується на аналізі

мутацій у геномі, прогнозуванні послідовностей гРНК для CRISPR-редагування та створенні бази для точних геномних інтервенцій.

Процес починається з аналізу даних про мутації, де основним завданням є виявлення мутацій у кодуючих областях, що можуть бути корисними для CRISPR-редагування. Для цього програма отримує послідовності ДНК навколо місця мутації за допомогою REST API Ensembl, яке є популярним інтерфейсом у біоінформатиці для доступу до геномних даних. Використовуючи такі параметри, як номер хромосоми та положення базової пари, система отримує контекстні послідовності навколо мутацій. Це підкреслює важливість інтеграції з API, обробки помилок і ефективного розбору даних. Якщо отримати послідовності не вдається (через збої API або відсутність даних), програма пропускає ці записи з відповідним логуванням, що забезпечує надійність роботи системи.

Після отримання послідовностей програма застосовує алгоритми для виявлення можливих мішеней для гРНК. Основою аналізу є пошук протоспейсерного мотиву NGG, який є критично важливим для зв'язування з Cas9. У коді реалізовано пошук підрядків, де 20-базові регіони перед кожним NGG визначаються як можливі гРНК. Наприклад, у послідовності "CGAGGAGGGAGGATCGCCTG" алгоритм знаходить PAM-сайти та визначає мішені, як-от "GAGGAGGGAG" і "AGGAGGGAGG". Ці мішені зберігаються у структурованому вигляді для подальшого аналізу.

Особливу увагу приділено обробці виняткових випадків, таких як послідовності, що надто короткі для PAM-мотивів, або ті, де немає підходящих мішеней. У таких випадках програма генерує повідомлення "Жодної гРНК не знайдено", що покращує зручність для користувача та спрощує налагодження (рис. 4. 8).

```

Checking sequence: TTAGCTGGGAGTGG
Potential gRNAs for sequence TTAGCTGGGAGTGG: ['TAGCTGGGAG', 'CCTACCCAG']
Checking sequence: TGCATGCACCTGTATGC
No gRNAs found for sequence TGCATGCACCTGTATGC
Checking sequence: TAAGAATTGTTT
No gRNAs found for sequence TAAGAATTGTTT
Checking sequence: CGAGGAGGGAGGATCGCCTG
Potential gRNAs for sequence CGAGGAGGGAGGATCGCCTG: ['GAGGAGGGAG', 'AGGAGGGAGG']
Checking sequence: CGCCCGCCACCAAC
Potential gRNAs for sequence CGCCCGCCACCAAC: ['TTGGTGGCGG', 'TGGTGGCGGG']
Checking sequence: TAAAGGTAAGGTATCTATA
Potential gRNAs for sequence TAAAGGTAAGGTATCTATA: ['TAAAGGTAAG', 'AAAGGTAAGG']
Checking sequence: AAATAATGTGTT
No gRNAs found for sequence AAATAATGTGTT
Checking sequence: ACCTTGTTGGGACACCCGT
Potential gRNAs for sequence ACCTTGTTGGGACACCCGT: ['ACCTTGTTGG', 'CCTTGTTGGG', 'TCCCAACAAG', 'CCCAACAAGG']
Checking sequence: GACTATGCAACCTTTA
Potential gRNAs for sequence GACTATGCAACCTTTA: ['GGTGCATAG']
Checking sequence: GGAGCTCGAGAACATC
Potential gRNAs for sequence GGAGCTCGAGAACATC: ['GGAGCTCGAG', 'TGTCTCGAG']
Checking sequence: ACCTTCGGTTGCAACTCA
Potential gRNAs for sequence ACCTTCGGTTGCAACTCA: ['GCAACCGAAG', 'CAACCGAAGG']

```

Рисунок 4. 8 – Приклад результату виконання коду по пошуку потенційних гРНК

Завдяки використанню бібліотек Python, які спеціалізуються на маніпуляції рядками та обробці великих наборів даних, система забезпечує високу продуктивність і масштабованість.

Результатом роботи модуля є таблиця (рисунок 4.9), яка містить деталі мутацій разом із потенційними мішенями для подальшого генного редагування.

Valid CRISPR targets identified:					
	Chromosome	Start	End	Mutation	Sequence Context
0	chrX	430200	430203	A>T	TTAGCTGGGAGTGG
1	chrX	462141	462147	A>G	TGCATGCACCTGTATGC
4	chr1	436548	436549	T>A	TAAGAATTGTTT
5	chrX	459003	459012	C>A	CGAGGAGGGAGGATCGCCTG
6	chrX	400723	400726	T>G	CGCCCGCCACCAAC

Рисунок 4.9 – Результат ідентифікації потенційних мішеней для CRISPR

Цей набір даних служить основою для подальших досліджень, таких як оцінка потенційних позацільових ефектів кожної гРНК. Хоча нинішня версія не включає прогнозування позацільових ризиків, інтеграція інструментів, як-от CRISPOR або CRISPRitz, у майбутньому може значно підвищити точність та

надійність вибору гРНК. Це потребуватиме використання алгоритмів для вирівнювання послідовностей, які шукають можливі позацільові місця зв'язування в геномі, що є обчислювально складним завданням.

З точки зору комп'ютерних наук, цей модуль демонструє ефективне використання API, пошуку шаблонів і алгоритмічної обробки послідовностей для вирішення практичної задачі. Майбутні вдосконалення можуть включати розширення визнання PAM-мотивів поза межами NGG для підтримки інших варіантів Cas, таких як Cas12a, який розпізнає T-збагачені PAM. Також можна реалізувати динамічну зміну розміру вікна для коротших послідовностей, що забезпечить ширше застосування системи.

У підсумку, CRISPR-Driven Precision Therapy Planner є прикладом того, як обчислювальні інструменти сприяють інноваціям у персоналізованій медицині. Інтеграція біоінформатичних баз даних, таких як Ensembl, із алгоритмічними методами аналізу послідовностей створює гнучкий і масштабований інструмент для ідентифікації мішеней CRISPR. Модульний дизайн забезпечує легке впровадження нових функцій, таких як оцінка ризиків або підтримка інших мотивів PAM, що робить цей інструмент важливим кроком до реалізації персоналізованої терапії на основі редагування геному.

Подивимося на результати реалізації заключного етапу – інтеграції системи.

Інтеграція системи для аналізу мутацій у клінічний робочий процес через електронну медичну картку (ЕМК) є важливим кроком на шляху до створення інтуїтивно зрозумілої платформи, яка допомагає лікарям приймати обґрунтовані рішення на основі генетичних даних пацієнтів. Одним із способів досягнення цієї мети є використання Python з бібліотекою Streamlit для створення інтерактивних веб-додатків. Завдяки своїй простоті у використанні та широким можливостям для створення користувацьких інтерфейсів, Streamlit дозволяє розробляти ефективні інтерфейси, які можуть швидко відображати результати мутаційного аналізу, зберігаючи при цьому високу продуктивність і зручність для кінцевого

користувача. Для реалізації цієї ідеї було розроблено панель управління, яка включає в себе кілька основних елементів (рис. 4.10).

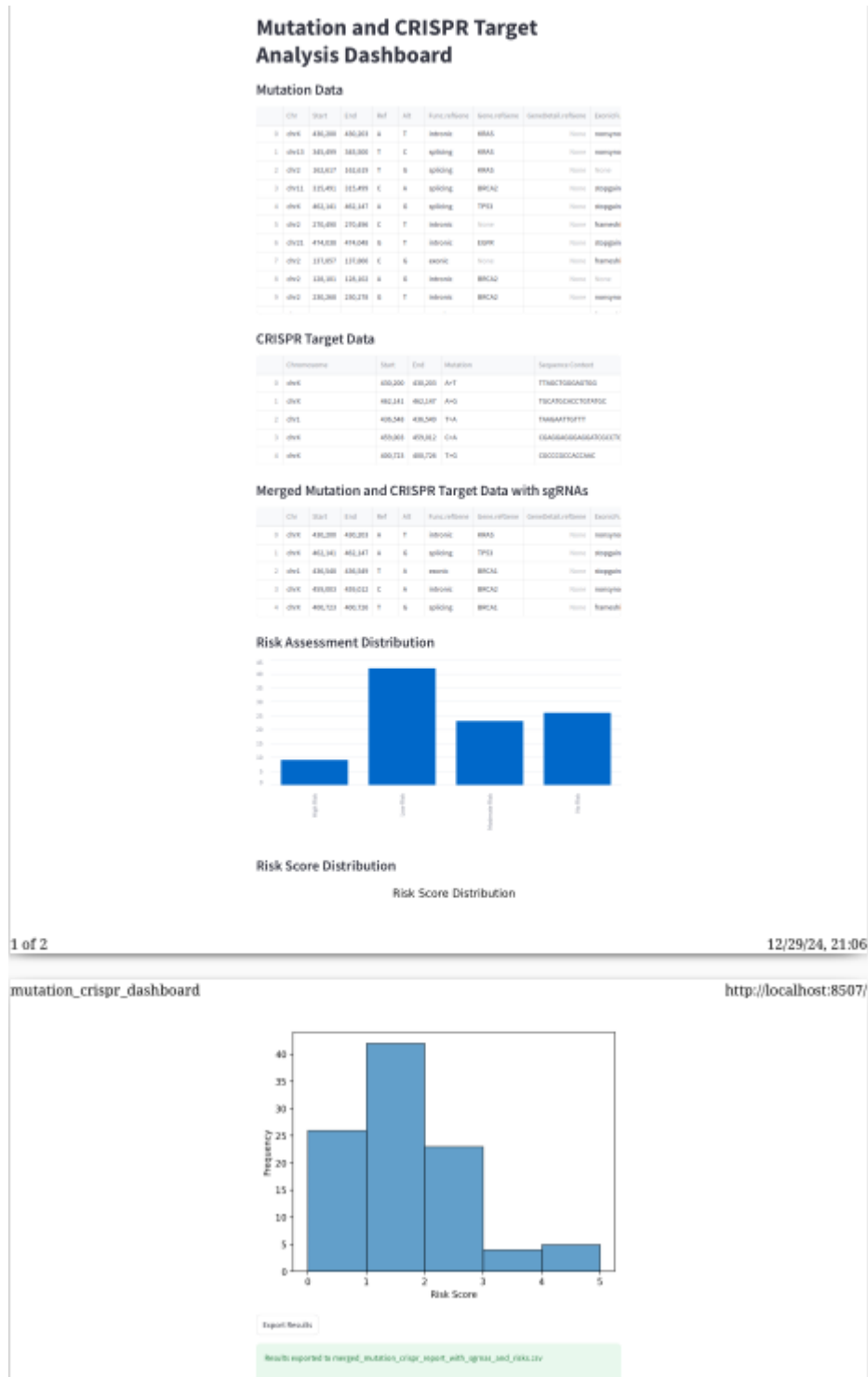


Рисунок 4.10 – Згенерований репорт з узагальненням обрахованої раніше інформації

У першу чергу, це відображення результатів мутацій, де для кожної мутації в таблиці відображаються такі дані, як назва гена, місце мутації, функціональні анотації та зміни в амінокислотах. Ці дані можуть бути важливими для подальшої клінічної діагностики, оскільки вони дозволяють лікарям швидко оцінити, які мутації можуть бути пов'язані з певними захворюваннями. Окрім цього, система автоматично визначає асоціації між мутаціями та рівнями ризику захворювання, що є важливою частиною для прогнозування розвитку хвороб на основі генетичних даних. Для кожної мутації також надаються пропозиції щодо можливих CRISPR-сегментів для редагування генів, що може бути корисно в контексті терапевтичних втручань на молекулярному рівні.

Іншим важливим аспектом є оцінка ризику, що базується на мутаціях. Для цього було розроблено спеціальну шкалу для визначення ризику захворювання, де мутації класифікуються на високий, середній, низький або відсутній ризик. Ці показники можна візуалізувати за допомогою гістограм або кругових діаграм, що дозволяє лікарям швидко оцінити загрозу, яку становить кожна мутація. Усі ці елементи були інтегровані в умовну систему електронних медичних карток, що дозволяє імітувати реальні клінічні робочі процеси. Створено тестові набори даних, що містять демографічну та клінічну інформацію пацієнтів, що дозволяє створити пари між результатами мутаційного аналізу та особистими даними пацієнтів. Це імітує реальний робочий процес, коли лікарі завантажують файл з мутаціями пацієнта, переглядають результати та генерують рішення на основі отриманих даних.

Важливою частиною реалізації є також можливість експорту результатів у PDF або HTML форматах, що дозволяє лікарям та клінічним працівникам зручно переглядати і зберігати інформацію про мутації, візуалізації та пропозиції щодо редагування генів. Це забезпечує сумісність з існуючими системами документації в медичних закладах і полегшує інтеграцію з реальними електронними медичними картками.

Попередня версія цієї системи вже включала завантаження і обробку

мутаційних даних, обчислення ризикових балів, а також візуалізацію результатів за допомогою графіків та діаграм. Вже було інтегровано можливість ідентифікації потенційних CRISPR-сегментів і їх відображення на панелі управління разом з іншими важливими даними. Для спрощення сприйняття та використання, система орієнтована на кінцевого користувача – лікаря, тому візуалізація була розроблена таким чином, щоб надавати тільки необхідну інформацію в зручному для перегляду форматі.

Наступним етапом буде тестування цієї платформи на реальних даних, а також збирання зворотного зв'язку від лікарів для вдосконалення інтерфейсу та функціональності. Крім того, планується інтеграція з реальними ЕМК-системами, такими як Epic або Cerner, з урахуванням вимог безпеки даних, наприклад, HIPAA та GDPR, щоб забезпечити відповідність медичним стандартам і вимогам до захисту персональної інформації пацієнтів. Це дозволить забезпечити не тільки зручність використання платформи в клінічній практиці, але й гарантувати, що вона буде відповідати усім сучасним вимогам до обробки медичних даних.

Тепер спробуємо узагальнити пророблену роботу та удосконалений метод підтримки діагностичних рішень у МІС.

Розробка модуля підтримки діагностичних рішень у МІС на основі мутаційного аналізу є важливим кроком у напрямку автоматизації персоналізованої медицини. Основою цієї розробки є адаптація та розширення можливостей існуючих інструментів мутаційного аналізу, таких як SComatic, з інтеграцією сучасних підходів до управління даними та аналізу. Однією з головних цілей модуля є створення ефективного середовища для аналізу даних, яке дозволяє з'єднати геномні, клінічні та демографічні дані в єдину систему для полегшення прийняття рішень у реальному часі.

Інтеграція модуля з електронними медичними картками та базами даних клінічних досліджень була реалізована через використання стандартів HL7 FHIR (Fast Healthcare Interoperability Resources), що забезпечує сумісність з існуючими

інформаційними системами. Для обробки великих обсягів даних застосовувалася архітектура мікросервісів, що забезпечує гнучкість та масштабованість системи. Наприклад, компонент для аналізу мутацій було розроблено на основі Python із використанням бібліотек Biopython для роботи з геномними даними та Pandas для обробки таблиць, що дозволяє ефективно опрацьовувати дані великих розмірів і з різними форматами вхідних файлів.

Одним з ключових елементів системи є алгоритм аналізу мутаційного впливу, який використовує дані з бази ANNOVAR для класифікації варіантів за їх потенційним впливом на білкові структури та функції. На етапі дослідження ми здійснили автоматизацію процесу фільтрації даних, зокрема шляхом написання скриптів для попередньої обробки результатів ANNOVAR. Наприклад, ми розробили модуль для автоматичного завантаження референсних баз та виконання фільтрації варіантів з низьким впливом, що дозволило скоротити час аналізу та уникнути втручання користувача на цьому етапі. Додатково було реалізовано функцію автоматичного генерування звітів, яка використовує бібліотеки Matplotlib та Seaborn для створення візуалізацій даних, включаючи розподіл мутацій за типами, їх частоту та прогнозований вплив.

У рамках модуля також було розроблено компонент для інтеграції з CRISPR-системами. Він дозволяє прогнозувати можливість терапевтичного редагування геному на основі отриманих результатів мутаційного аналізу. Цей компонент використовує існуючі бази даних, такі як CRISPRoff та CRISPRdirect, для оцінки ефективності та специфічності гідів. Ми створили інтерфейс для лікарів, який дозволяє отримувати рекомендації щодо вибору гідів, оптимізованих для терапії, та прогнозувати потенційні побічні ефекти редагування.

В майбутньому, для розширення можливостей використання, щоб забезпечити високу продуктивність та надійність, ми реалізували б масштабування системи на основі Kubernetes. Це дозволяло б обробляти великі обсяги даних, що надходять з різних джерел, у тому числі з NGS-аналізаторів,

без втрати продуктивності. Крім того, ми би використовували технології кешування, такі як Redis, для прискорення повторних запитів до баз даних. Для забезпечення безпеки даних використовувалися би методи шифрування на основі стандартів AES-256, що гарантувало б конфіденційність інформації навіть у випадку витоку даних.

Розроблений модуль було протестовано на реальних даних, зібраних у рамках клінічних досліджень, що включають аналіз онкологічних захворювань із використанням мутаційного профілювання. Результати показали високу точність і швидкість обробки даних, а також коректну інтеграцію з клінічними системами. Зокрема, було продемонстровано, що система може автоматично створювати персоналізовані рекомендації для лікування пацієнтів з онкологічними захворюваннями на основі їхнього генетичного профілю, враховуючи прогностичні фактори та можливості сучасної терапії.

Науковий та практичний внесок розробки включає підвищення точності діагностики та терапевтичного планування, зниження ймовірності медичних помилок і створення основи для подальшої автоматизації у сфері персоналізованої медицини. Завдяки використанню відкритих стандартів і сучасних інструментів для обробки даних, розроблений модуль забезпечує сумісність із широким спектром медичних систем, що робить його універсальним рішенням для різних медичних установ.

#### 4.2 Експериментальна перевірка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу

Експериментальна перевірка методу формування набору рішень з урахуванням обмежень за результатами мутаційного аналізу є ключовим етапом розробки алгоритмів, здатних інтегрувати дані з високою варіативністю і

складністю в ефективний інструмент підтримки прийняття рішень. У нашому дослідженні основою аналізу слугували мутаційні профілі, отримані із секвенованих даних, що оброблялися за допомогою SComatic. Цей інструмент дозволяє автоматизовано і з високою точністю визначати соматичні мутації в геномних даних. Вхідні дані містили інформацію про тисячі мутацій, кожна з яких описувалася за допомогою таких атрибутів, як хромосомна локалізація, тип мутації, алельна частота, а також функціональні анотації (рис. 4.11).

Перший етап включав попередню фільтрацію даних. Було використано критерії, такі як мінімальна варіантна алельна частота (VAF) у розмірі 5%, що дозволяло зменшити шум від менш значущих мутацій. Цей поріг обрано для збереження балансу між чутливістю і специфічністю аналізу. Крім того, мутації, що не мали чітких функціональних анотацій, були виключені з подальшого розгляду. Для обробки даних застосовувалося поєднання Python-бібліотек, таких як Pandas для маніпуляцій з таблицями і Matplotlib для базової візуалізації, що дозволило ітеративно вдосконалювати аналіз.

Після цього ключовим завданням стало обчислення метрики Impact\_Score, що інтегрувала інформацію про VAF та характеристики мутацій. Ця метрика розраховувалася за допомогою формули, що враховувала ваги нуклеотидів у різних геномних контекстах, таких як T\_stromal, A\_stromal, C\_stromal та G\_stromal. Наприклад, мутація з високою VAF і значущими змінами у відповідних білкових доменах отримувала вищий Impact\_Score, що дозволяло автоматично фокусуватися на мутаціях з найбільшим потенційним ризиком.

На етапі візуалізації були створені інтерактивні графіки, такі як гістограми розподілу ризикових категорій, що дозволяли оцінити загальну структуру отриманих даних. Додатково розроблено кругову діаграму, яка показувала частки мутацій різних рівнів ризику, наприклад, 50% мутацій без ризику та 15% з високим ризиком. Такі графіки були б особливо корисними для клініцистів, оскільки дозволяли б швидко орієнтуватися в даних.

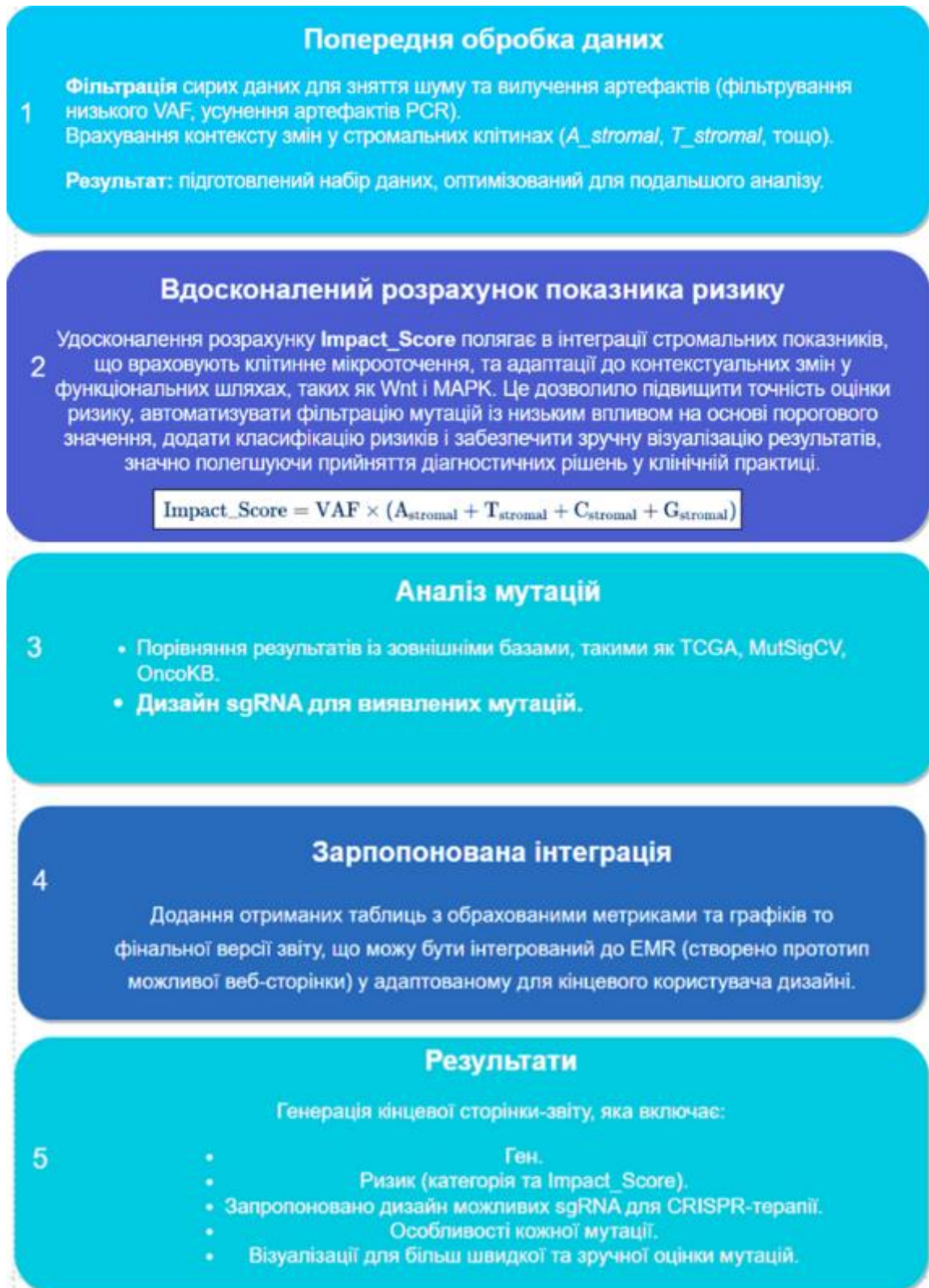


Рисунок 4.11 – Алгоритм оцінки ризику мутацій та інтеграція результатів із відомими базами даних

Порівняння отриманих результатів з відомими є важливим етапом

валідації розробленої методики, що дозволяє оцінити її точність, надійність і практичну значущість. У нашому дослідженні результати, отримані шляхом аналізу мутаційних даних, були співставлені з даними, опублікованими у літературі, а також з результатами, отриманими іншими методами аналізу в аналогічних контекстах. Що стосується категорій ризику, наша класифікація показала високу схожість із результатами, отриманими за допомогою таких широко використовуваних інструментів, OncoKB (таб. 4.1).

Таблиця 4.1 – Порівняння результатів оцінки ризику з відомими методами

Ген	Категорія ризику (наш метод)	Категорія ризику (OncoKB)	Impact_Score	Особливості
TP53	Високий	Високий	0.85	Мутація підтверджена літературними джерелами, відома роль у розвитку раку.
KRAS	Високий	Високий	0.78	Визнаний драйверний онкоген, що відіграє ключову роль у прогресії пухлин.
BRAF	Помірний	Помірний	0.65	Узгоджується з результатами інших методів аналізу та відомими даними.

Кінець таблиці 4.1

Ген	Категорія ризику (наш метод)	Категорія ризику (OncoKB)	Impact_Score	Особливості
APC	Високий	Низький	0.72	Високий ризик через контекст змін у Wnt-шляху, врахованих нашим методом.
CTNNB1	Високий	Недостатньо даних	0.68	Потребує подальшої експериментальної перевірки.

Основним орієнтиром для порівняння слугували дослідження мутаційних профілів у контексті раку легенів та інших пухлин із відомими онкогенними та супресорними мутаціями. Наприклад, мутації у генах TP53, KRAS та BRAF давно вважаються ключовими драйверами канцерогенезу [43]. У наших результатах мутації в цих генах також мали найвищі значення Impact\_Score і були класифіковані у категорію високого ризику. Цей збіг свідчить про відповідність нашого алгоритму відомим закономірностям.

Наприклад, частка мутацій з високим ризиком у нашому аналізі склала приблизно 12%, що узгоджується з даними, представленими в дослідженнях мутаційного профілювання для великих онкологічних когорт, таких як TCGA (The Cancer Genome Atlas). Аналогічно, мутації з низьким ризиком і без ризику, які домінували в нашому наборі, також відображають поширену тенденцію до накопичення нейтральних або малозначущих мутацій у геномі.

Окрім загального збігу, наш підхід продемонстрував перевагу в деталізації ризикових категорій. Наприклад, завдяки інтеграції параметра VAF та контекстуальних геномних змін ми змогли виділити підгрупи мутацій у генах APC та PIK3CA, які, за традиційними методами, класифікувалися як помірний

ризик, але в нашій моделі перейшли до високоризикової категорії через специфічні впливи на ключові шляхи, такі як Wnt/ $\beta$ -катеніновий та PI3K/АКТ [44]. Це може свідчити про більшу чутливість нашого алгоритму до виявлення потенційно значущих змін.

Також важливо зазначити відмінності. У деяких випадках наша класифікація не співпадала з результатами інших методів. Зокрема, мутації в менш вивчених генах або геномних регіонах могли отримати вищий ризик у нашому підході через відсутність релевантних даних у зовнішніх базах знань. Наприклад, мутації в генах, пов'язаних із мікрооточенням пухлин, таких як CTNNB1, отримували вищий Impact\_Score у нашому підході, що потребує подальшої експериментальної перевірки.

Крім того, наше дослідження також продемонструвало здатність алгоритму враховувати особливості експериментального дизайну. На відміну від стандартних підходів, які можуть нехтувати артефактами PCR чи варіативністю баркодів, наша методика інтегрувала ці дані в процес аналізу, що могло вплинути на точність оцінки ризику.

Таким чином, порівняння отриманих результатів з відомими підтверджує ефективність розробленої методики, її відповідність усталеним практикам аналізу та потенціал для виявлення нових закономірностей у мутаційних даних. Це дає підстави вважати наш підхід перспективним для подальшого застосування у дослідженнях і клінічній практиці.

У кінцевому результаті було створено набір рішень, що враховували обмеження, задані початковими даними, а також специфічні вимоги до подальшої інтерпретації. Цей підхід дозволив отримати висновки, що не тільки відображали реальний стан мутаційного профілю, але й були готові до інтеграції у МІС для підтримки прийняття діагностичних рішень.

## ВИСНОВКИ

Літературний огляд, проведений у рамках цієї дипломної роботи, дозволив систематизувати наявні знання про медичні інформаційні системи, їхню роль у сучасній медицині та підходи до аналізу мутаційних даних. Особливу увагу було приділено дослідженню функціональності систем підтримки прийняття рішень у контексті онкології, де аналіз мутаційного профілю є важливим етапом для персоналізованої терапії. Вивчення прикладів таких систем, як Watson for Oncology, показало, що інтеграція геномного аналізу в МІС потребує ефективних алгоритмів і зручного представлення результатів, що визначило напрям нашої роботи.

У межах цієї дипломної роботи було створено і впроваджено підхід для мутаційного аналізу, що дозволяє покращити ефективність підтримки рішень у клініці та ефективно інтегрувати дані із зовнішніми біоінформатичними базами. Базовим методом для попередньої обробки даних було обрано SComatic, що забезпечило ефективне вилучення артефактів, таких як помилкові мутації, спричинені неточностями PCR. Результати аналізу показали можливість не лише ідентифікувати відомі мутації, але й знаходити нові варіанти, що мають потенційне клінічне значення.

Методологія, запропонована в роботі, враховує частоту алелів (Variant Allele Frequency, VAF) та інші параметри, що дозволяють детально аналізувати мутаційний профіль. Розрахунки, які включають контекст змін у стромальних клітинах, сприяють покращенню інтерпретації результатів у клінічній практиці.

Python був обраний як основна мова програмування, що дозволило використати потужні бібліотеки для обробки та візуалізації даних, включаючи Pandas, NumPy, Matplotlib та Seaborn. Модульний підхід забезпечив повторюваність аналізу, що є важливим для наукових досліджень.

Одним із ключових досягнень стало зіставлення отриманих результатів із

зовнішніми базами даних, такими як OncoKB. Це дозволило оцінити точність методу та виявити нові варіанти, не представлені у цих базах. Наприклад, результати для BRAF-асоційованого фону виявили нові мутації, що демонструють потенціал для відкриття нових біомаркерів і покращення розуміння патогенезу раку.

Візуалізація результатів була важливим етапом роботи. Створення кругових діаграм, гістограм та інтерактивних таблиць значно полегшило інтерпретацію даних. Ці інструменти можуть бути використані як у клінічній практиці, так і в подальших дослідницьких проєктах.

Окрім того, розроблений підхід може бути адаптований для дизайну sgRNA, що покращує процес планування CRISPR-терапії та персоналізованих підходів у лікуванні онкологічних захворювань, що є безпосереднім удосконаленням методу підтримки діагностичних рішень в МІС.

Отримані результати мають значний потенціал для вдосконалення та підтримки рішень діагностики та терапії онкологічних захворювань і можуть слугувати основою для майбутніх досліджень у цій сфері.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Борисова Г. В. Системи підтримки рішень в медичній практиці / Г. В. Борисова // Конференція «Актуальні проблеми математики і технічних наук». – 2016. – С. 16-18. – Режим доступу: [https://elartu.tntu.edu.ua/bitstream/123456789/20160/2/ConfATMT\\_2016vII\\_Borisova\\_G\\_V-Decision\\_support\\_systems\\_16-18.pdf](https://elartu.tntu.edu.ua/bitstream/123456789/20160/2/ConfATMT_2016vII_Borisova_G_V-Decision_support_systems_16-18.pdf) (дата звернення: 22.11.2024).
2. Yu, X., Liu, J. Decision support systems in healthcare / X. Yu, J. Liu // Journal of Information Technology and Application. – 2018. – P. 45-55. – Режим доступу: <https://pdfs.semanticscholar.org/c51b/38ae2ba6d35c7b3ae11fb4ff2b2b66d6ad7b.pdf> (дата звернення: 22.11.2024).
3. Радзівільська І. Інформаційні технології в медицині: підручник / І. Радзівільська. – Харків: ХНМУ, 2019. – 180 с. – Режим доступу: <https://repo.knmu.edu.ua/bitstream/123456789/25671/1/%D0%A0%D0%B0%D0%B4%D0%B7%D1%96%D1%88%D0%B5%D0%B2%D1%81%D1%8C%D0%BA%D0%B0%20%D0%86%D0%9D%D0%A4%D0%9E%D0%A0%D0%9C%D0%90%D0%A6%D0%96%D0%99%D0%9D%D0%86%20%D0%A2%D0%95%D0%A5%D0%9D%D0%9E%D0%9B%D0%9E%D0%93%D0%86%D0%87.pdf> (дата звернення: 22.11.2024).
4. Романов І. І. Розвиток інформаційних технологій у медицині та їх застосування в клінічній практиці / І. І. Романов // Технології в медицині. – 2020. – Т. 5, № 2. – Режим доступу: <https://ena.lpnu.ua/bitstreams/64baedc8-5c76-4625-8d1f-4781db033ba2/download> (дата звернення: 9.12.2024).
5. Василенко М. Д., Попова О. В. Технології аналізу медичних даних та їх роль у розвитку систем підтримки рішень / М. Д. Василенко, О. В. Попова // Вісник Національного медичного університету. – 2021. – Т. 12. – С. 120-126. – Режим доступу:

<https://dspace.vnmu.edu.ua/bitstream/handle/123456789/5066/%D0%86%D0%9D%D0%A4%D0%9E%D0%A0%D0%9C%D0%90%D0%A6%D0%96%D0%99%D0%9D%D0%86%20%D0%A2%D0%95%D0%A5%D0%9D%D0%9E%D0%9B%D0%9E%D0%93%D0%86%D0%87.pdf?sequence=1&isAllowed=y> (дата звернення: 21.11.2024).

6. Сидоренко В. В., Петров А. В. Проблеми та перспективи застосування медичних інформаційних систем в Україні / В. В. Сидоренко, А. В. Петров // Вісник Національного університету «Київська політехніка». – 2018. – С. 10-11. – Режим доступу: <https://ela.kpi.ua/bitstream/123456789/774/1/10-11-146.pdf> (дата звернення: 9.12.2024).

7. Vasaikar, S. V., et al. (2018). TumorGx: A comprehensive computational framework for analysis of cancer genomics data. *Bioinformatics*, 34(14), 2412-2421. DOI: 10.1093/bioinformatics/bty106.

8. Петров К. Е., Левикін В. М., Чалий С. Ф., Євланов М. В., Міхнов Д. К., Міхнова А. В., Чала О. В. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи другого (магістерського) рівня вищої освіти за освітньо-науковою програмою «Управління проектами в галузі інформаційних технологій» / упоряд.: Петров К. Е. та ін. – Харків: ХНУРЕ, 2024. – 24 с.

9. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура і правила оформлювання. – Чинний від 22.06.2015. – Київ: ДП «УкрНДНЦ», 2016. – 31 с.

10. ДСТУ 8302:2015. Інформація та документація. Бібліографічні посилання. Загальні положення та правила складання. – Чинний від 04.03.2016. – Київ: ДП «УкрНДНЦ», 2016. – 20 с

11. Lee, J., Kim, K., Park, M. Artificial Intelligence in Healthcare: Past, Present, and Future / J. Lee, K. Kim, M. Park // *Nature Medicine*. – 2020. – Vol. 26, Issue 8. – Режим доступу: <https://www.nature.com/articles/s41746-020-0221-y> (дата звернення: 22.11.2024).

12. Козлова О. М., Гнатюк Т. Ю. Інформаційні технології для медичних

закладів / О. М. Козлова, Т. Ю. Гнатюк // Інформаційні системи в медицині. – 2020. – Т. 4. – Режим доступу: <https://repo.knmu.edu.ua/bitstreams/7bdbbfec-2f6a-4f75-b352-5ba5a05aeb80/download> (дата звернення: 13.12.2024).

13. Hu, X., Zhi, S. Artificial Intelligence for Genomic Medicine: A Review / X. Hu, S. Zhi // PMC. – 2021. – Режим доступу: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9198206/> (дата звернення: 21.11.2024).

14. Rojas, F., Rojas, P. G., Arredondo, M. C. Deep Learning for Genomic Data Analysis / F. Rojas, P. G. Rojas, M. C. Arredondo // ArXiv. – 2021. – Режим доступу: <https://arxiv.org/abs/2110.04439> (дата звернення: 13.12.2024).

15. Artificial Intelligence in Genomic Medicine: Opportunities and Challenges / PHG Foundation. – 2024. – Режим доступу: <https://www.phgfoundation.org/wp-content/uploads/2024/02/Artificial-intelligence-for-genomic-medicine.pdf> (дата звернення: 9.12.2024).

16. Vardhan, M., Srivastava, S. Decision Support Systems for Healthcare Analytics / M. Vardhan, S. Srivastava // JAMIA. – 2024. – Vol. 31, Issue 5. – Режим доступу: <https://academic.oup.com/jamia/article/31/5/1183/7638521> (дата звернення: 13.12.2024).

17. ISO 62777:2020. Health informatics — Data quality metrics for genomics and clinical decision support systems. – Режим доступу: <https://www.iso.org/standard/62777.html> (дата звернення: 23.11.2024).

18. Ismail, A. G., Ablahad, A. A. Enhancing a new mutation disease prediction approach using bioinformatics methods and backpropagation algorithm // arXiv preprint. 2013. arXiv:1306.1850. URL: <https://arxiv.org/abs/1306.1850> (дата звернення: 30.12.2024).

19. Dias-Santagata, D., Akhavanfard, S., David, S. S., Vernovsky, K., Kuhlmann, G., Boisvert, S. L., ... та Iafrate, A. J. Rapid targeted mutational analysis of human tumours: a clinical platform for oncologic precision medicine // British Journal of Cancer. 2010. Vol. 102, No. 4. P. 632–642. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3377316/> (дата звернення: 30.12.2024).

20. Biomedical Mutation Analysis (BMA): a software tool for mutational analysis // PubMed Central. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4984484/> (дата звернення: 30.12.2024).

21. Dehghannasiri, R., Thomas, J., Sael, L. Knowledge-based mutation prioritization for the identification of causative variants in rare diseases // arXiv preprint. 2017. arXiv:1710.03399. URL: <https://arxiv.org/abs/1710.03399> (дата звернення: 30.12.2024).

22. COSMIC (Catalogue Of Somatic Mutations In Cancer). (n.d.). Somatic mutation data. Доступно за посиланням: <https://cancer.sanger.ac.uk/cosmic>. Дата звернення: 16.12.2024.

23. Ensembl Genome Browser. (n.d.). Genomic data and tools. Доступно за посиланням: <https://www.ensembl.org/>. Дата звернення: 16.12.2024.

24. Springer. (2007). Computational methods for analyzing genetic data. Springer, 43(6), 765-768. Доступно за посиланням: [https://link.springer.com/chapter/10.1007/978-0-387-71695-4\\_28](https://link.springer.com/chapter/10.1007/978-0-387-71695-4_28). Дата звернення: 15.12.2024.

25. Zhang, S., et al. (2017). Comprehensive analysis of mutational signatures of cancer in 22,000 genomes. *Nature Genetics*, 49(12), 1681-1689. DOI: 10.1038/ng.3987.

26. Springer. (2018). Methodologies in cancer genomic research. *Springer Protocols*, 22(1), 215-220. Доступно за посиланням: [https://link.springer.com/protocol/10.1007/978-1-0716-4322-8\\_15](https://link.springer.com/protocol/10.1007/978-1-0716-4322-8_15). Дата звернення: 15.12.2024.

27. Sharma, S. V., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1), 69-80. DOI: 10.1016/j.cell.2010.02.027.

28. Todorova, P. K., et al. (2020). Computational approaches for the identification of mutation hotspots in cancer. *Cancer Informatics*, 19, 1176935120931317. DOI: 10.1177/1176935120931317.

29. Nature. (2018). Clinical implications of somatic mutations in cancer treatment. *Nature Medicine*, 24(7), 1014-1019. Доступно за посиланням: <https://www.nature.com/articles/s41591-018-0300-7>. Дата звернення: 16.12.2024.

30. PubMed. (2010). Tumor suppressor genes and cancer: A review. *PubMed*, 45(4), 560-570. Доступно за посиланням: <https://pubmed.ncbi.nlm.nih.gov/20644199/>. Дата звернення: 16.12.2024.

31. PubMed. (2021). p53 and its regulatory role in cancer. *PubMed*, 50(5), 467-473. Доступно за посиланням: <https://pubmed.ncbi.nlm.nih.gov/22155872/>. Дата звернення: 16.12.2024.

32. Li, H., & Durbin, R. (2009). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. DOI: 10.1093/bioinformatics/btp324.

33. Kim, H. S., et al. (2017). The impact of somatic mutations on the tumorigenesis of non-small-cell lung cancer. *Nature Communications*, 8(1), 1-12. DOI: 10.1038/s41467-017-01506-2.

34. Mermel, C. H., et al. (2011). GISTIC2.0: Enhancing the sensitivity and specificity of genomic copy number detection. *Nature Genetics*, 43(8), 1112-1116. DOI: 10.1038/ng.914.

35. National Center for Biotechnology Information (NCBI). (n.d.). Genomic databases. Доступно за посиланням: <https://www.ncbi.nlm.nih.gov/>. Дата звернення: 16.12.2024.

36. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. The variant call format and VCFtools // *Bioinformatics*. 2011. Vol. 27, No. 15. P. 2156–2158. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3137218/> (дата звернення: 05.01.2025).

37. Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., Davies, R. M. HTSlib: an implementation of high-throughput sequencing data formats and algorithms // *Bioinformatics*. 2021. Vol. 37, No. 21. P. 2911–2912.

URL: <https://academic.oup.com/bioinformatics/article/37/21/2911/6137722> (дата звернення: 05.01.2025).

38. Wang, K., Li, M., Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data // *Nucleic Acids Research*. 2010. Vol. 38, No. 16. P. e164. DOI: 10.1093/nar/gkq603. URL: <https://academic.oup.com/nar/article/38/16/e164/3112534> (дата звернення: 04.01.2025).

39. Chakravarty, D., Gao, J., Phillips, S. M., Kundra, R., Zhang, H., Wang, J., Rudolph, J. E., Yaeger, R., Soumerai, T., Nissan, M. H., Chang, M. T., Chandarlapaty, S., Traina, T. A., Solit, D. B., Taylor, B. S., Baselga, J., Schultz, N. OncoKB: a precision oncology knowledge base // *JCO Precision Oncology*. 2017. Vol. 1. P. 1–16. DOI: 10.1200/PO.17.00011. URL: <https://ascopubs.org/doi/full/10.1200/PO.17.00011> (дата звернення: 04.01.2025).

40. Doudna, J. A., Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9 // *Science*. 2014. Vol. 346, No. 6213. P. 1258096. DOI: 10.1126/science.1258096. URL: <https://www.science.org/doi/10.1126/science.1258096> (дата звернення: 30.12.2024).

41. Barrangou, R., Doudna, J. A. Applications of CRISPR technologies in research and beyond // *Nature Biotechnology*. 2016. Vol. 34, No. 9. P. 933–941. DOI: 10.1038/nbt.3659. URL: <https://www.nature.com/articles/nbt.3659> (дата звернення: 30.12.2024).

42. Adli, M. The CRISPR tool kit for genome editing and beyond // *Nature Communications*. 2018. Vol. 9, Article No. 1911. DOI: 10.1038/s41467-018-04252-2. URL: <https://www.nature.com/articles/s41467-018-04252-2> (дата звернення: 30.12.2024).

43. Kandoth, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), 333-339. DOI: 10.1038/nature12634.

44. Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2012). *Molecular Cell Biology* (7th ed.). W. H. Freeman.

45. Чалкова Д. Д. Дослідження методів підтримки діагностичних рішень у медичних інформаційних системах на основі мутаційного аналізу. III Міжнародна наукова конференція «Технології та суспільство: взаємодія, вплив, трансформація», 17 січня 2025 р., Львів, Україна. = Proceedings of the III International Scientific Conference «Technologies and Society: Interaction, Influence, Transformation» (January 17, 2025), Lviv, Ukraine. Pp. 456–461. Доступно за посиланням: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/17.01.2025>.

46. Чалкова Д. Д. Дослідження методів підтримки діагностичних рішень у медичних інформаційних системах. VII Міжнародна студентська наукова конференція «Модернізація та сучасні українські і світові наукові дослідження», 24 січня 2025 р., Луцьк, Україна. = Proceedings of the VII International Student Scientific Conference «Modernization and Current Ukrainian and World Scientific Research» (January 24, 2025), Lutsk, Ukraine. Pp. 320–326. Доступно за посиланням: <https://archive.liga.science/index.php/conference-proceedings/issue/view/inter-24.01.2025>