

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: Комплексний мета-аналіз глобальних наборів даних про COVID-19

Виконав: студент IV курсу, групи СН-41

спеціальності 122 Комп'ютерні науки

(шифр і назва спеціальності)

(підпис)

Князь Ю.І.

(прізвище та ініціали)

Керівник

(підпис)

Пасічник В.В.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Шимчук Г.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Гащин Н.Б.

(прізвище та ініціали)

Тернопіль
2021

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Боднарчук І.О.
(підпис) (прізвище та ініціали)

«__» _____ 2021 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Бакалавр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Князь Юрій Ігорович
(прізвище, ім'я, по батькові)

1. Тема роботи Комплексний мета-аналіз глобальних наборів даних про COVID-19

Керівник роботи Пасічник Володимир Володимирович, д.т.н., професор кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «02» березня 2021 року № 4/7-171

2. Термін подання студентом завершеної роботи 21 червня 2021р.

3. Вихідні дані до роботи Наукові публікації та вільнодоступні колекції даних щодо COVID-19

4. Зміст роботи (перелік питань, які потрібно розробити)
Вступ. 1 COVID-19, джерела даних та їх конвеєрне опрацювання. 2 Комплексний метааналіз глобальних наборів даних щодо COVID-19. 3. Безпека життєдіяльності, основи хорони праці. Висновки. Перелік джерел. Додатки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Титульний аркуш. 2. Тема, мета та задачі дослідження. 3. Актуальність досліджень.

4. Сутності предметної області COVID-19. 5. Основні етапи конвеєрного дослідження даних щодо COVID-19. 6. Генерування та аналітичне опрацювання даних Omics-даних.

7. Процес інтеграції даних для формування графічного подання інтерактом COVID-19.

8. Процес перепризначення лікарських препаратів. 9. Методи аналітичного опрацювання відомостей щодо COVID-19. 10. Аналітичне опрацювання відомостей щодо COVID-19.

11. Висновки. 12. Доповідь завершено.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи охорони праці	Гурик О.Я., доцент кафедри МТ		

7. Дата видачі завдання 7 червня 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	07.06.2021	<i>Виконано</i>
2.	Підбір джерел про набори даних щодо COVID-19, методи та засоби аналітичного опрацювання.	08.06.2021-09.06.2021	<i>Виконано</i>
3.	Переклад та опрацювання джерел про набори даних щодо COVID-19, методи та засоби їх аналітичного опрацювання.	10.06.2021-11.06.2021	<i>Виконано</i>
4.	Виконання дослідження щодо комплексного метааналізу глобальних наборів даних про COVID-19	12.06.2021-13.06.2021	<i>Виконано</i>
5.	Оформлення розділу «COVID-19, джерела даних та їх конвеєрне опрацювання»	14.06.2021-15.06.2021	<i>Виконано</i>
6.	Оформлення розділу «Комплексний метааналіз глобальних наборів даних щодо COVID-19»	16.06.2021-17.06.2021	<i>Виконано</i>
7.	Виконання завдання до підрозділу «Безпека» життєдіяльності	17.06.2021	<i>Виконано</i>
8.	Виконання завдання до підрозділу «Основи охорони праці»	17.06.2021	<i>Виконано</i>
9.	Оформлення кваліфікаційної роботи	18.06.2021	<i>Виконано</i>
10.	Нормоконтроль	19.06.2021	<i>Виконано</i>
11.	Перевірка на плагіат	19.06.2021	<i>Виконано</i>
12.	Попередній захист кваліфікаційної роботи	19.06.2021	<i>Виконано</i>
13.	Захист кваліфікаційної роботи	21.06.2021	

Студент

_____ (підпис)

Князь Ю.І.

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Пасічник В.В.

_____ (прізвище та ініціали)

АНОТАЦІЯ

Комплексний мета-аналіз глобальних наборів даних про COVID-19 // Кваліфікаційна робота освітнього рівня «Бакалавр» // Князь Юрій Ігорович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СН-41 // Тернопіль, 2021 // С. 49, рис. – 11, табл. – 6, кресл. – 12, додат. – 5, бібліогр. – 64.

Ключові слова: covid-19, аналітика, дані, класифікація, метод, структура, опрацювання.

Кваліфікаційна робота присвячена комплексному аналізу структури, вмісту, методів та засобів аналітичного опрацювання глобальних наборів даних про COVID-19. Мета роботи: підвищення рівня поінформованості громадян та медичних працівників щодо COVID-19.

В першому розділі кваліфікаційної роботи Проведено аналіз предметної області. Описано джерела даних про COVID-19, подано їх інтегральну характеристику та конвєсне опрацювання даних.

В другому розділі кваліфікаційної роботи описано аналітичне опрацювання даних SARS-CoV-2 на основі відомостей БД Omics. Розглянуто філогенний та мутаційний варіаційний аналіз даних щодо COVID-19. Досліджено аналітичне опрацювання генетичної інформації щодо COVID-19. Подано особливості інтерактного аналізу відомостей щодо COVID-19. Висвітлено аналіз зображень грудної клітки для діагностики та моніторингу COVID-19. Описано аналітичне опрацювання в процесах використання та перепризначення лікарських препаратів для COVID-19. Зокрема, наведено класифікаційну таблицю методів аналітичного опрацювання відомостей щодо COVID-19.

ANNOTATION

Complex meta-analysis of global datasets about COVID-19 // Qualification work of the educational level "Bachelor" // Kniaz Yuriy Ihorovych // Ternopil Ivan Pulyuy National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Sciences, group SN-41 // Ternopil, 2021 // P. 49, fig. – 11, tables – 6, chair. – 12, annexes – 5, ref. – 64.

Key words: covid-19, analytics, data, classification, method, structure, processing.

Qualification work is devoted to a comprehensive analysis of the structure, content, methods and tools of analytical processing of global datasets about COVID-19. Purpose: to raise awareness of citizens and health professionals about COVID-19.

In the first section of the qualification work the analysis of the subject area is carried out. The sources of data about COVID-19 are described, their integral characteristic and conveyor data processing are given.

The second section of the qualification work describes the analytical processing of SARS-CoV-2 data based on information from the Omics database. Phylogenetic and mutational variational analysis of data on COVID-19 are considered. Analytical processing of genetic information according to COVID-19 was investigated. Features of interactive analysis of information on COVID-19 are presented. Analysis of chest images for diagnosis and monitoring of COVID-19 is covered. Analytical processing in the processes of use and reassignment of drugs for COVID-19 is described. In particular, the classification table of methods of analytical processing of information on COVID-19 is given.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

DNN (англ. Deep neural network) – Глибинна нейронна мережа.

FDA (англ. U.S. Food and Drug Administration) – Управління з продовольства і медикаментів США.

GAN (англ. Generative Adversarial Network) – Генеративна конкурентна мережа.

GCN (англ. Global Cycling Network) – Глобальні циклічні мережі.

CNN (англ. Convolutional Neural Network) – Згорткові нейронні мережі.

LSTM (англ. Long Short-Term Memory) – Довга короткотривала пам'ять.

SNP (англ. Single nucleotide polymorphism) – Однонуклеотидний поліморфізм.

БД – База даних.

ВООЗ – Всесвітня організація охорони здоров'я.

КТ – Комп'ютерна томографія.

РНК – Рибонуклеїнова кислота.

ШІ – Штучний інтелект.

ЗМІСТ

ВСТУП		7
1 COVID-19, ДЖЕРЕЛА ДАНИХ ТА ЇХ КОНВЕЄРНЕ ОПРАЦЮВАННЯ.....		9
1.1 Аналіз предметної області.....		9
1.2 Джерела даних щодо COVID-19.....		11
1.2.1 Дані Omics.....		12
1.2.2 Інтерактомні дані		13
1.2.3 Дані рентенографічних зображень грудної клітки		14
1.2.4 Епідеміологічні дані.....		15
1.2.5 Бази даних медичних препаратів.....		17
1.3 Конвеєрне опрацювання даних щодо COVID-19.....		18
1.4 Висновок до першого розділу		22
2 КОМПЛЕКСНИЙ МЕТААНАЛІЗ ГЛОБАЛЬНИХ НАБОРІВ ДАНИХ ЩОДО COVID-19.....		23
2.1 Метааналіз глобальних наборів даних щодо COVID-19.....		23
2.2 Аналітичне опрацювання даних SARS-CoV-2 Omics		23
2.3 Філогенний та мутаційний варіаційний аналіз даних щодо COVID-19		24
2.4 Аналітичне опрацювання генетичної інформації щодо COVID- 19.....		25
2.5 Інтерактомний аналіз відомостей щодо COVID-19.....		26
2.6 Аналіз зображень грудної клітки для діагностики та моніторингу COVID-19		27
2.7 Аналітичне опрацювання в процесах використання та перепризначення лікарських препаратів для COVID-19.....		29
2.8 Аналітичне опрацювання метаболічних змін для важких та фатальних випадків COVID-19		32

	6
2.9 Висновок до другого розділу	37
3 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ХОРОНИ ПРАЦІ	38
3.1 Актуальність безпеки життєдіяльності людини.....	38
3.2 Долікарська допомога при пораненнях.....	40
3.3 Висновок до третього розділу	41
ВИСНОВКИ.....	42
ПЕРЕЛІК ДЖЕРЕЛ	43

ВСТУП

Актуальність теми. Новий коронавірус SARS-CoV-2, який зазвичай називають COVID-19 поширився у всьому світі з осені 2019 року та спричинив глобальну пандемію. COVID-19 викликає важке запалення легень та органів дихання, яке може призвести до летального результату.

На даний час невпинно зростає кількість досліджень щодо діагностування захворювання, його перебігу та наслідків. Спостерігається масове і швидке зростання досліджень, спрямованих на виявлення механізмів зараження COVID-19 та прогресування захворювання. Мета-аналіз цих наборів та колекцій даних має великий нерокритий потенціал для аналітичного опрацювання з метою підвищення рівня інформування щодо клінічної медичної допомоги та сприяння розвитку терапевтичних засобів.

Наявність великих за обсягом мультиомічних, радіологічних, біомолекулярних та медичних наборів медичних даних щодо COVID-19 вимагає розроблення нових дослідницьких та прогнозних моделей або адаптації існуючих.

Велика кількість підходів потребує напрацювання множини методів, засобів та інструментів для аналітичного опрацювання відомостей щодо COVID-19. Тому комплексний метааналіз глобальних наборів даних щодо COVID-19 є актуальним напрямком сучасних досліджень в галузі інформаційних технологій.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня «Бакалавр» є підвищення рівня поінформованості пацієнтів та медичних працівників щодо COVID-19. Для досягнення поставленої мети потребують вирішення ряд наступних завдань:

- Проаналізувати стан досліджень в галузі аналітичного опрацювання відомостей щодо COVID-19.
- Класифікувати та описати джерела даних щодо COVID-19.

– На основі відомих наукових публікацій провести комплексний метааналіз глобальних наборів даних щодо COVID-19, проаналізувати методи та засоби їх аналітичного опрацювання.

Практичне значення одержаних результатів, полягає в тому, що:

– Прокласифіковано та проаналізовано джерела даних щодо COVID-19.

– Описано ключові етапи конвеєрного опрацювання відомостей щодо COVID-19.

– На основі аналізу обширного переліку вітчизняних та закордонних публікацій сформовано описи поширених методів та засобів аналітичного опрацювання відомостей щодо COVID-19.

– Сформовано таблицю методів аналітичного опрацювання відомостей щодо COVID-19 в залежності від множини типів даних.

1 COVID-19, ДЖЕРЕЛА ДАНИХ ТА ЇХ КОНВЕЄРНЕ ОПРАЦЮВАННЯ

1.1 Аналіз предметної області

На даний час у світі відбувається масовий спалах гострої респіраторної коронавірусної інфекції SARS-CoV-2 [1] спричинений небезпечним для людського життя патогенним захворюванням, яку ВООЗ назвала COVID-19 [1]. Надзвичайно швидка передача інфекції від людини до людини спричинила експоненціальне збільшення кількості випадків захворювання за відносно малий період часу [2].

Багато дослідників та фахівців світової наукової спільноти з різних галузей зосередили свої зусилля на розробці вакцин, клінічній терапії, управлінні пацієнтами та ресурсами для боротьби з вірусом COVID-19 [3]. Як наслідок, зростає кількість відповідних вільнодоступних наборів даних. Зокрема, БД GISAID зібрала понад сямдесят тисяч вірусних геномних ланцюжків, а «Інформаційна панель Джона Хопкінса» [4] на даний час є одним з основних джерел даних для моніторингу епідеміологічних захворювань. Експоненційно-швидке зростання обсягів колекцій та наборів даних про лабораторні дослідження щодо COVID-19 потребує значних обчислювальних ресурсів та зусиль для практичної реалізації аналітичних підходів, сформованих на обчисленнях, зокрема глибинного навчання, штучного інтелекту [5]. Це, в свою чергу призводить до накопичення великих за обсягом колекцій та наборів даних, алгоритмів, програмного забезпечення та інструментів, які потребують відповідної організації та класифікації. На даний час основними складовими досліджуваної предметної області є (див. рисунок 1.1):

- джерела даних;
- сховища даних;
- моделі даних;

- методи та засоби для забезпечення процесів прийняття рішень;
- засоби інтерпретації.

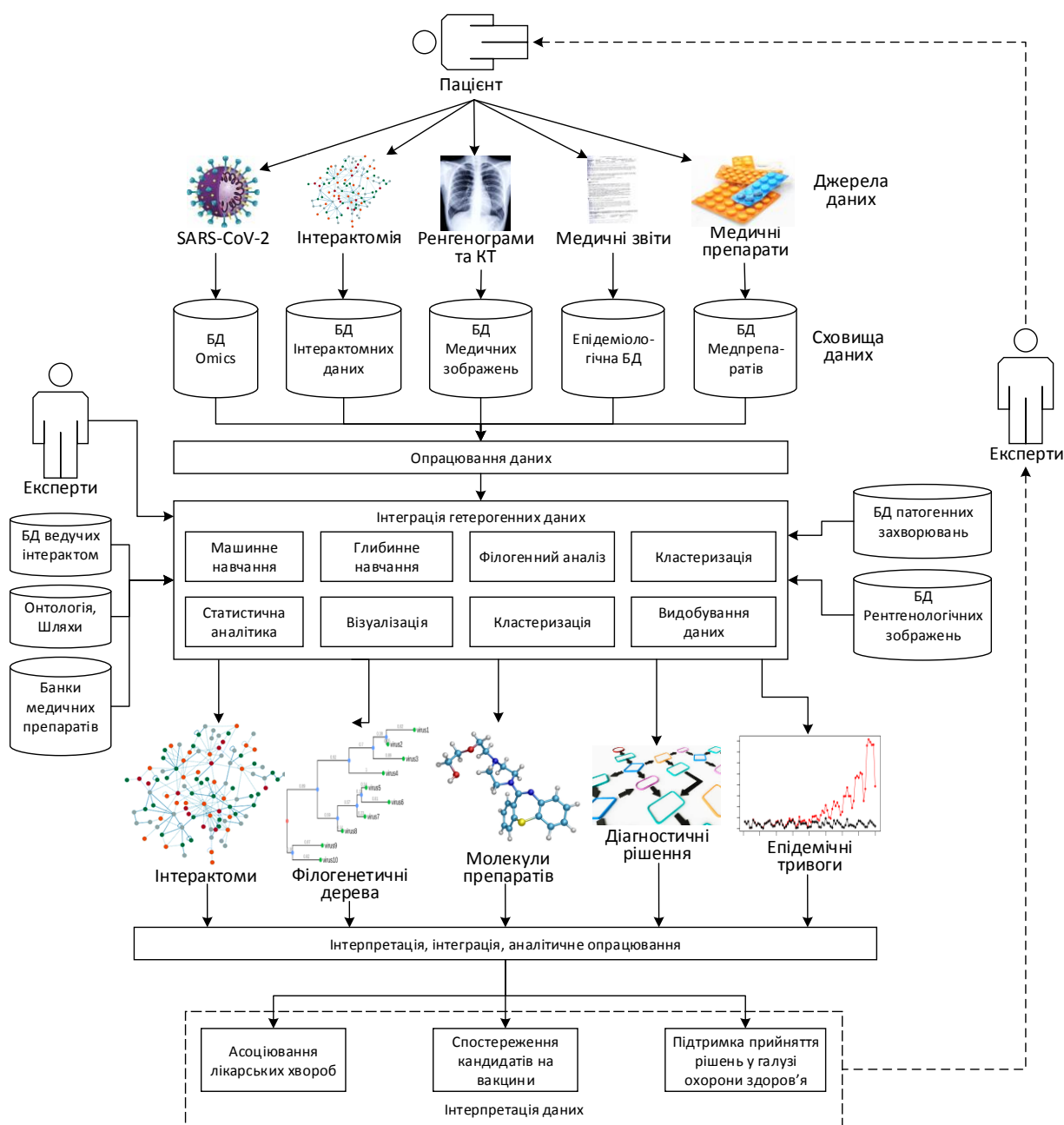


Рисунок 1.1 – Структура сутностей предметної області COVID-19 [6]

На рис. 1.1 подано процеси аналітичного опрацювання даних щодо COVID-19. Деякі з них формують нові знання, дані або моделі, які можуть бути використані в якості вхідних наборів при реалізації зворотного зв'язку в робочих процесах аналізу даних.

1.2 Джерела даних щодо COVID-19

Пандемія COVID-19 спричинила безпрецедентно-масове та швидке зростання обсягів збирання та зберігання відомостей та наукових публікацій по всьому світі [7]. На рисунку 1.2 в графічній формі подано зростання тенденцій пов'язаних із COVID-19 наукових публікації. Для пошуку відомостей використовувались «Dimensions» та «Google Scholar» [8]. Пошук відбувався по ключових словах «COVID-19», «COVID-19 Машинне навчання» та «COVID 19 наука про дані» з використанням пошукової фільтрації опублікованих статей та передруків, книг та монографій, матеріалів та розділів.

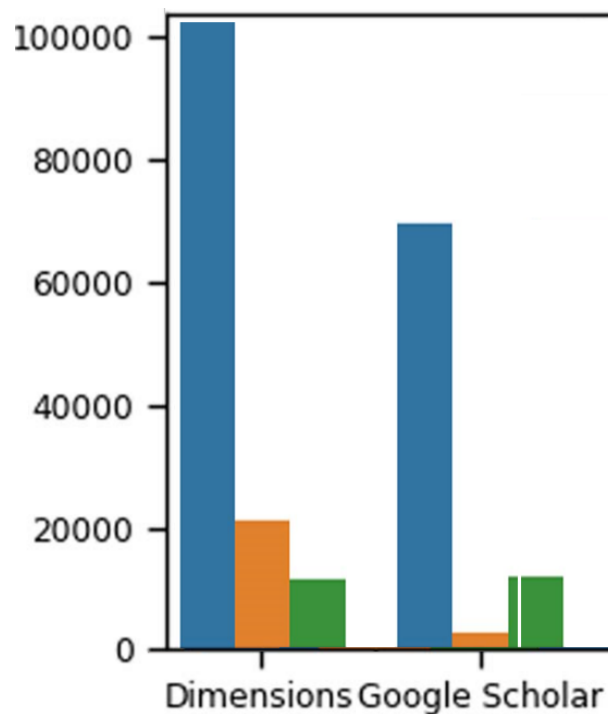


Рисунок 1.2 – зростання тенденцій пов'язаних із COVID-19 наукових публікації

За відомостями академічної пошукової системи Dimension [9] на даний час у відкритому доступі присутні понад сімсот тридцять наборів даних щодо COVID-19. Зокрема опубліковано первинні дані, зокрема послідовності,

клінічні зображення, медичні звіти та вторинні дані, зокрема, білкові структури, інтерактоми, епідеміологічні відомості. Первинні дані безпосередньо генеруються безпосередньо на основі вірусу або пацієнта. На їх основі формуються уточнені, узагальнені та виведені результуючі набори – вторинні дані. В контексті COVID-19 процеси аналізу даних та формування висновків використовують еталонні набори для видобування або виведення нових знань.

1.2.1 Дані Omics

Високопродуктивні технології Omics [10] використовують високоефективні аналітичні процедури для оперативного виявлення та кількісної оцінки клітинних процесів з метою виявлення певних молекул у біологічних зразках зібраних з використанням біохімічних аналізів. Дані Omics зазвичай відносяться до класів геноміки, транскриптоміки, протеоміки та метаболоміки. Технології та дані Omics використовуються для покращення розуміння характеристик та особливостей нових малодосліджених або невідомих вірусів. Вони використовувались для секвенування геному SARS-CoV-2 для з'ясування особливостей росту, мутації та реплікування COVID-19 [11]. В таблиці 1.1 подано популярні сховища даних Omics для наукових досліджень щодо COVID-19.

Таблиця 1.1 – Сховища даних Omics для наукових досліджень щодо COVID-19 [6]

Тип даних	Репозиторій	Опис	Джерело
1	2	3	4
Нуклеотиди та білки	GISAID	Понад 75 000 вірусних геномних послідовностей SARS-CoV-2	https://www.gisaid.org/

Продовження таблиці 1.1

1	2	3	4
Нуклеотиди та білки	NCBI	Більше 25 000 нуклеотидів, 250 401 білки	https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/
Структура	RCSB PDB	Білки SARS-CoV-2	https://www.rcsb.org/covid19
Структура	SWISS-MODEL	Білки SARS-CoV-2	https://swissmodel.expasy.org/repository/species/2697049
Неоднорідні	COVID-19 hg		https://www.covid19hg.org
Метаболоміка	iProX	Інтегрований ресурсний центр протеомів	https://www.iprox.org/
Транскриптоміка	NCBI	Дані експресії РНК	http://www.ncbi.nlm.nih.gov/geo/
Транскриптоміка	OmicsDI	Дані експресії РНК	https://www.omicsdi.org/

Відкриття важливих біомаркерів COVID-19 відбулося завдяки дослідженню РНК SARS-CoV-2 на основі відомостей NCBI [12] та OmicsDI [13]. Аналітичне опрацювання метаболічних даних може використовуватись у визначенні потенційних хімічних біомаркерів для COVID-19. Розділ метаболоміки використовується для здійснення аналізу фенотипів та отримання знань щодо метаболічних станів біологічних систем. На даний час опубліковано досить мало робіт щодо метаболічних зразків COVID-19.

1.2.2 Інтерактомні дані

Інтерактомія – це вивчення особливостей біохімічної взаємодії між біологічними молекулами, зокрема, білками, факторами транскрипції, та малими молекулами. Саме ці види взаємодії є елементарними сутностями для пошуку кореляції та відношень в БД. В таблиці 1.2 подано популярні сховища інтерактомних даних для наукових досліджень щодо COVID-19.

Подана в [14] гомологія SARS-CoV2 та групи інших коронавірусів використана для формування прогнозованих взаємодій між вірусними білками та білками типу хазяїн-вірус.

Таблиця 1.2 – Сховища інтерактомних даних для наукових досліджень щодо COVID-19 [6]

Тип даних	Репозиторій	Опис	Джерело
Взаємодія, мережа	BioGRID	Понад 800 взаємодіючих білків	https://thebiogrid.org/
Взаємодія, графік	IntAct	4479 парні взаємодії	https://www.ebi.ac.uk/intact/
Взаємодіючий білок	HPA	Понад 200 взаємодіючих людських білків з SARS-CoV-2	https://www.proteinatlas.org/humanproteome/SARS-CoV-2

«Мокрий» лабораторний підхід [15] було застосовано для клонованих протеом SARSCoV-2. Автори [16] подають опис діагностичних інструментів сформованих на основі набору даних про взаємодію SARSCoV-2, який було зібрано консорціумом «IMEx».

1.2.3 Дані рентгенографічних зображень грудної клітки

Технології рентгенографії та КТ можуть бути використані для діагностування уражень легенів та дихальних шляхів, спричинених COVID-19. На даний час зображення «GGO» є найпоширенішим серед КТ-зображень грудної клітки інфікованих COVID-19 пацієнтів. Характерними є мультифокальні, периферичні та двосторонні зображення. В таблиці 1.3 подано популярні сховища рентгенографічних зображень для наукових досліджень щодо COVID-19.

Під час COVID19 «GGO» може виглядати як однофокальне ураження, найчастіше розташоване в нижній частці правої легені [17], але вони можуть використовуватись для спостереження прогресу захворювання. Нагальна потреба у формуванні автоматичних діагностичних інструментів для швидкого виявлення інфікованих COVID-19 громадян стимулює наукову

спільноту розробляти нові діагностичні засади на основі інформаційних технологій машинного навчання.

Таблиця 1.3 – Сховища рентгенографічних зображень для наукових досліджень щодо COVID-19 [6]

Тип даних	Репозиторій	Опис	Джерело
Рентгенографія	github	Понад 800 зображень	https://github.com/ieee8023/covid-chestxray-dataset
КТ	github	349 зображень щодо 216 пацієнтів	https://github.com/UCSD-AI4H/COVID-CT
	github	63 849 зображень щодо 377 пацієнтів	https://github.com/mr7495/COVID-CTset
	github	104 009 КТ-зображень 1489 пацієнтів	https://github.com/lindawangg/COVID-Net/
	MosMED	КТ грудної клітки з висновками щодо COVID-19	https://mosmed.ai/en/
Рентгенографія та КТ	BIMCV-COVID19+	Рентгенівські зображення CXR (CR, DX), 1380 CX, 885 DX та 163 CT	https://osf.io/nh7g8/

Зокрема інформаційно-технологічна платформа «TrainingData.io» [18] пропонує безкоштовний інструмент обміну анотаціями навчальних даних, котрі можуть бути використані для формування нових моделей машинного навчання.

1.2.4 Епідеміологічні дані

Епідеміологічні дані – це множина результатів неекспериментальних спостережень, отриманих внаслідок накопичування відомостей з будь-якого джерела даних, котре пов'язане з галуззю охорони здоров'я та експертами

медичних доменів. Іноді вони включають екологічні, клінічні та лабораторні дані, статистичні результати географічного поширення тощо. На даний час багато незалежних дослідницьких груп та організацій збирають та зберігають епідеміологічні дані. Група авторів на чолі з Донг [19] спроектували та розробили інформаційну панель для забезпечення цілодобового безкоштовного доступу до даних в галузі охорони здоров'я. При цьому зібрано практично з усього світу пов'язані із зареєстрованими випадками COVID-19 дані, зокрема відомості щодо інфікування, смерті та одужання пацієнтів. Уряд Італії запровадив реалізацію подібного інформаційного ресурсу з вихідними даними щодо COVID-19 [20]. Ху [21] розробив базу даних з відкритим доступом для зберігання інформації щодо лабораторних обстежень пацієнтів. В таблиці 1.4 подано популярні сховища епідеміологічних даних для наукових досліджень щодо COVID-19.

Таблиця 1.4 – Сховища епідеміологічних даних для наукових досліджень щодо COVID-19 [6]

Тип даних	Репозиторій	Опис	Джерело
Інформація	CIDRAP	Зареєстровані випадки коронавірусної хвороби, звіт про ситуацію, епідеміологія, вірусологія, клінічні особливості	https://www.cidrap.umn.edu/COVID-19~/epidemiology
	WHO	Інформація щодо COVID-19	https://covid19.who.int/
	Italian Civil Protection		https://github.com/pcm-dpc/COVID-19-
	SCIENTIFIC DATA	Вибрані індивідуальні дані з національних, провінційних та муніципальних звітів про стан здоров'я та Інтернет-звітів	https://doi.org/10.6084/m9.figshare.11974344

Всі подані проекти мають ряд спільних характеристик:

- використання простих форматів даних;
- процедури експорту у стандартний формат обміну даними;
- прості інтерфейси формування та опрацювання запитів;

- інтеграція геолокаційних даних;
- містять демографічні відомості.

В багатьох вільнодоступних наборах та колекціях збережено відомості щодо пересуваннями громадян, симптомів та демографічних показників.

1.2.5 Бази даних медичних препаратів

Медичний препарат – це невелика молекула органічного походження [22], яка може здійснювати функції активації або інгібування терапевтично значущих білків в процесі захворювання. Процеси відкриття нових препаратів та схвалення «Адміністрацією з харчових продуктів та медикаментів» (FDA) на даний час є доволі складними, фінансово витратними та трудомісткими. На першому етапі відбувається ідентифікація препарату та цілі, а на другому відбувається розвиток малих молекул, здатних взаємодіяти з ціллю [23]. Схвалені молекулярні формули ліків та описи мішеней зазвичай зберігаються у загальнодоступних базах даних, здебільшого у спрощеній формі молекулярного введення (SMILE) для забезпечення можливості переназначення ліків або комерційної розробки. В таблиці 1.5 подано описи популярних бази даних медичних препаратів для наукових досліджень щодо COVID-19.

Таблиця 1.5 – Бази даних медичних препаратів для наукових досліджень щодо COVID-19 [6]

Тип даних	Репозиторій	Опис	Джерело
1	2	3	4
Молекула	Drugbank	Містить відомості щодо близько 13 607 медичних препаратів	https://www.drugbank.ca/ COVID-19
	PubChem	Найбільша база даних у світі: понад 350 мільйонів сполук, речовин, BioAssay	https://pubchem.ncbi.nlm.nih.gov/

Продовження таблиці 1.5

1	2	3	4
	ChEMBL	Біоактивні молекули, пов'язані з SARS-CoV-2, із властивостями, подібними до медичних препаратів	https://www.ebi.ac.uk/chembl/
	Excelra	Ліки, пов'язані з COVID-19, які перебувають на «клінічній, доклінічній та експериментальній» стадії	https://www.excelra.com/COVID-19-drug-repurposing-database/
	CAS	Противірусні препарати та супутні хімічні сполуки для лікування захворювання COVID-19	https://www.cas.org/
	Pharmaceutical	Препарати на всіх стадіях доклінічного та клінічного розвитку за показами щодо COVID-19	https://www.pharmaceutical-technology.com/coronavirus-drug-trials-studies/

Зокрема, DrugBank – це сховище відомостей щодо ліків та цілей, а його останній випуск містить 13 597 записів щодо медичних препаратів, у тому числі 2 641 опис схвалених молекул препаратів, 1 390 схвалених біологічних молекул білків, пептидів, вакцин та алергенів, 130 нутрицевтичних препаратів та понад 6 377 експериментальних препаратів, які знаходяться на стадії клінічних досліджень та 5 225 непотрібних білкових послідовностей у формі мішень-фермент-транспорт-носій.

1.3 Конверсне опрацювання даних щодо COVID-19

На даний час, «наука про дані» – це нова міждисциплінарна дослідницька галузь, яка використовує інноваційні методи та засоби, процеси та алгоритми, для видобування та одержання нових колекцій відомостей та знань із великих за обсягом та швидкоплинних наборів даних різноманітного

походження та структури. Войтек [24] стверджує, що ермін «наука про дані» був вперше використаний Патілом та Хамербахером в 2008р.

На рисунку 1.3 подано основні етапи конвеєрного опрацювання даних щодо COVID-19 з метою забезпечення процедур аналітичного дослідження та підтримки процесів прийняття рішень.

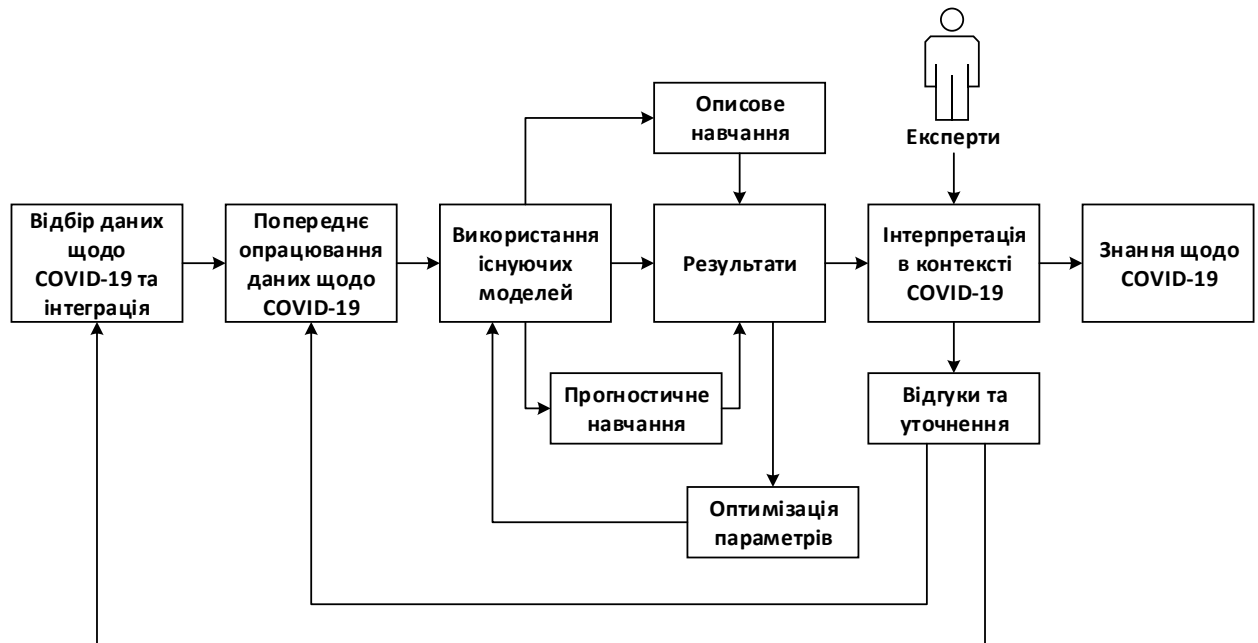


Рисунок 1.3 – Основні етапи конвеєрного дослідження даних щодо COVID-19

Конвеєрне опрацювання в основному складаються з чотирьох фаз:

- збір необроблених даних;
- попереднє опрацювання;
- описове або прогнозне моделювання;
- інтерпретація.

На першому етапі відбувається відбір та інтеграція даних щодо COVID-19 з обширної множини джерел. Потім відбувається їх попереднє опрацювання та фільтрація з метою усунення неінформативних або недостовірних значень, зокрема, таких що виходять за межі ранжування або зашумлені. На наступному етапі відбувається використання існуючих

моделей з метою розуміння структури та природи даних щодо COVID-19 або видобування важливих відповідних зразків, що описують дані або використовуються для прогнозування асоціативних залежностей. Зрештою повинна відбутись інтерпретація результатів та їх пояснення експертами доменів предметних областей, пов'язаних з COVID-19. При цьому на кожному з етапів аналізу можуть бути згенеровані виправлення або уточнення, які застосовуються до будь якого з попередніх етапів.

Інтеграція різнорідних даних щодо COVID-19 є вирішальним етапом «науки про дані». Успішне застосування будь-якої моделі «науки про дані» залежить від якості даних про COVID-19. Через недоліки структури та значень, шум або помилки у прокурах створення та відбору даних щодо COVID-19, результати робочого процесу з їх опрацювання даних можуть призвести до неправильних результатів та, як наслідок, помилкової та малоефективної інтерпретацій. Тому важливо коректно застосовувати різнотипові процедури очищення та фільтрації даних. Стандартизація та перетворення даних щодо COVID-19 необхідні при формуванні даних з різнотипових джерел. В комплексі всі вищеперелічені етапи формують процес попереднього опрацювання даних щодо COVID-19 [25]. На наступних етапах потрібно виконати процедури для аналітичного дослідження даних щодо COVID-19, яке виконують передавання даних до обраної обчислювальної моделі. Будь-який з процесів «науки про дані» повинен включати статистичний опис, зокрема тип даних щодо COVID-19, їх розподіл, суттєві особливості, взаємозв'язок між сутностями та значеннями вхідного набору даних. Це призводить до кращого розуміння самих даних щодо COVID-19, структури процедур та етапів, котрі використовуються для попереднього опрацювання та аналітичного опрацювання. Коректний вибір функціональних наборів допомагає визначити та сформулювати перелік відповідних атрибутів в наборі даних. Візуалізація відомостей щодо COVID-19 спрямована на вивчення можливих взаємозв'язків між функціями

наборів даних або між різними за своєю природою та походженням наборами даних. Розмірність та зменшення колекцій даних допомагають зробити процес опрацювання даних щодо COVID-19 більш ефективним та стійким до шумів, помилкових та недостовірних відомостей.

Колекції дані щодо COVID-19 можуть бути проаналізовані за допомогою описових, тобто неконтрольованих, та прогнозних, тобто контрольованих моделей, які часто можуть бути коректно та комплексно об'єднані між собою [26]. Слід відзначити ключові основи процедур глибинного навчання, які можуть бути корисними для аналітичного опрацювання прогнозних даних щодо COVID-19. Згорткові нейронні мережі (CNN) [27] ефективно використовуються для аналітичного опрацювання рентгенологічних зображень грудної клітки пацієнтів інфікованих COVID-19. Окремим видом CNN-моделі, який спеціально розробленою для опрацювання графічно поданих відомостей або мереж, є мережі із згортанням графіків (GCN) [28]. На даний час, через відсутність адекватних колекцій навчальних зразків даних щодо COVID-19 для коректного навчання глибинних моделей значну роль відіграють синтетичні набори даних [29]. Нещодавно було зроблено важливий прорив у архітектурі глибинних моделей, котрий відбувся завдяки генеративними конкурентним мережами (GAN) [30] для формування колекцій синтетичних даних подібних до реальних колекцій. Структура GAN сформована на основі двох одночасно навчених нейронних мереж – генератора та дискримінатора. Генератор формує фальшиві екземпляри а дискримінатор розпізнає згенеровані навчальні зразки. Це дозволяє одночасно покращувати обидва модулі в процесі вивчення важливих розрізнявальних функцій для вихідного набору відомостей щодо COVID-19. Регресійний аналіз – це контрольована модель аналітичного опрацювання, яка також може бути використана для прогнозування тенденцій пандемії COVID-19 [31]. Кластеризація [32] – це загально відома в наукових колах модель неконтрольованого машинного навчання, яка

визначає та узагальнює прихований шаблон всередині даних на основі визначених метрик близькості. Результати використання поданих вище методів аналітичного опрацювання відомостей щодо COVID-19 повинні перевіряти експерти відповідних доменів. Зворотній зв'язок експертів в галузях пов'язаних з COVID-19 може використовуватись для уточнення етапів конвеєрного опрацювання даних.

1.4 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр» подано аналіз предметної області. Описано джерела даних щодо COVID-19 та подано їх інтегральну характеристику. Висвітлено конвеєрне опрацювання даних щодо COVID-19.

2 КОМПЛЕКСНИЙ МЕТААНАЛІЗ ГЛОБАЛЬНИХ НАБОРІВ ДАНИХ ЩОДО COVID-19

2.1 Метааналіз глобальних наборів даних щодо COVID-19

Комплексний та багатофакторний мета-аналіз наявних на даний час наборів даних є перспективним підходом для формування безпосереднього та всебічного розуміння процесів інфікування, перебігу, лікування та постлікувальної реабілітації COVID-19 [33]. При цьому важливим завданням є виявлення однозначних та надійних біомаркерів для підвищення ефективності клінічної допомоги пацієнтам, розвитку діагностичних та терапевтичних засобів.

Проведення метааналізу наборів даних щодо COVID-19 дає можливість покращити розуміння загальних та унікальних генетичних моделей інфекції [34]. Однак, на даний час комплексний метааналіз наборів даних щодо COVID-19 доволі утруднений через велику кількість непідтверджених та неперевірених даних, практичні труднощі в узгодженні високих рівнів неоднорідності, властивих глобальним наборам даних щодо COVID-19. Щоб усунути ці недоліки та отримати краще та повніше розуміння змін, що лежать в основі захворювання COVID-19, потрібно провести комплексний та системний аналітичний огляд досліджень щодо опрацювання наборів даних про COVID-19 та сформувавши перелік етапів для конвеєрного опрацювання з використанням різнотипових методів та засобів аналітичного опрацювання.

2.2 Аналітичне опрацювання даних SARS-CoV-2 Omics

Метою дослідження колекцій даних Omics є оперативне виявлення проксимального походження COVID-19, ідентифікація його мутацій та формування прогнозних моделей для ефективною ідентифікації COVID-19.

Для визначення вірусного геному COVID-19 дослідники використовують нуклеотидні послідовності та прогнозування 3D-структури білка (див. рисунок 2.1).

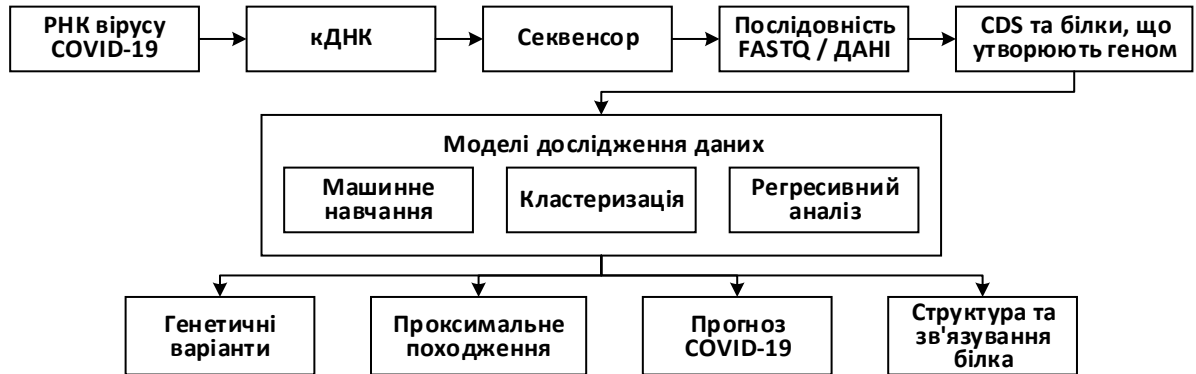


Рисунок 2.1 – Генерація та аналітичне опрацювання даних Omics-даних

Фрагменти послідовностей нуклеїнових кислот вірусу COVID-19, отримані з організмів інфікованих громадян, після попередньої підготовки та проведення лабораторних аналізів використовуються як вхідні дані для алгоритмів аналітичного опрацювання даних. Ці дослідження використовуються для:

- Аналізу генетичних варіантів COVID-19.
- Аналіз геномів вірусу COVID-19, що інфікують різні біологічні види.
- Прогнозування білкових взаємодій COVID-19.
- Збирання відомостей щодо структури та динаміки вірусних білків COVID-19.

2.3 Філогенний та мутаційний варіаційний аналіз даних щодо COVID-19

Виявлення еволюційного походження COVID-19 на даний час є найбільш актуальною темою досліджень обчирного кола науковців спільнот та колективів. Воно спрямоване на формування еволюційного дерева вірусу

COVID-19 на основі найближчих видів [35]. Для цього адаптовано методи кластеризації, котрі ефективно використовуються з метою створення філогенезу серед сімейства коронавірусів SARS-COV та MERS-COV [36], [37] до якої відноситься COVID-19. Еволюційні процеси вірусних інфекцій можуть бути описані на основі алгоритмів «*silico*» з використанням методів філогенетичного аналізу. Створення філогенетичного дерева в своїй основі опирається на аналіз РНК-последовності шляхом вирівнювання. На даний час для вирівнювання геному SARS-CoV-2 дослідниками використано невеликий перелік загальноприйнятих методів, зокрема «DNAMAN» [38], ClustalW [39], MUSCLE [40], Jalview [41] та MAFT [42]. На основі відомостей отриманих внаслідок вирівнювання РНК виводиться еволюційна історія за допомогою методу найближчих сусідів або максимальної ймовірності [43]. Для вірусів, що швидко розвиваються, зокрема SARS-CoV-2 [44], для ідентифікації однонуклеотидних поліморфізмів (SNP) здебільшого використовуються підходи до вирівнювання РНК або методи формування последовності на основі похідних ознак.

2.4 Аналітичне опрацювання генетичної інформації щодо COVID-19

Для з'ясування механізму зараження COVID-19 важливим є визначення повної последовності його геному та циркулюючих варіантів за допомогою моделей машинного навчання у комплексному поєднанні з порівняльною геномікою. Для формування навчальних колекцій даних використовуються методи вирівнювання РНК та без вирівнювання РНК. З метою підвищення ефективності моделей машинного навчання використовується множина ознак, зокрема *k*-mers – підпоследовності довжини *k*, та *N*-грами – последовності *N* елементів даного зразка, хімічні властивості амінокислот та відомості щодо мутації COVID-19, котрі видобуваються з використанням

методів вирівнювання. Останні наукові публікації спрямовані на поєднання п'яти загально відомих в наукових колах моделей класифікації у виборі ознак, отриманих на основі набору вірусних геномів, що належать до обширного сімейства коронавірусів та геномів SARS-CoV-2, для виявлення нових версій SARS-CoV-2. Фанг використовує для цього двосторонні згорткові нейронні мережі CNN [45].

2.5 Інтерактомний аналіз відомостей щодо COVID-19

Аналіз множинних інтерактом (взаємодій) – ще метод, який використовується для інтеграції даних, отриманих з гетерогенних білкових або генних мереж. На даний час інтерактомічні дослідження, пов'язані з SARS-CoV-2, переслідують такі основні цілі:

- розробка можливих методів лікування COVID-19;
- допомоги постраждалим від COVID-19 особам;
- введення нової вакцини для блокування поширення COVID-19.

Незважаючи на існування обширної множини лабораторій, що проводять дослідження щодо COVID-19 у різних галузях, продовж останнього періоду часу наукові публікації містять значну кількість прикладів використання мас-спектрометрії для визначення взаємодій білків SARS-CoV-2 [46]. Один з перших підходів щодо побудови інтерактому SARS-CoV-2 описаний в [47].

Автори [48] запропонували інтеграцію отриманих з доступних БД відомостей щодо РНК та даних про експресію генів. Структура процесу інтеграції даних для формування графічного подання інтерактом COVID-19 наведена на рисунку 2.2. Авторами [49] описано спробу шаблон кодону використовується для висновку про можливі взаємодії між 26 білками SARS-CoV-2 та селективними білками-господарями, що беруть участь у 17 основних сигнальних шляхах клітин.

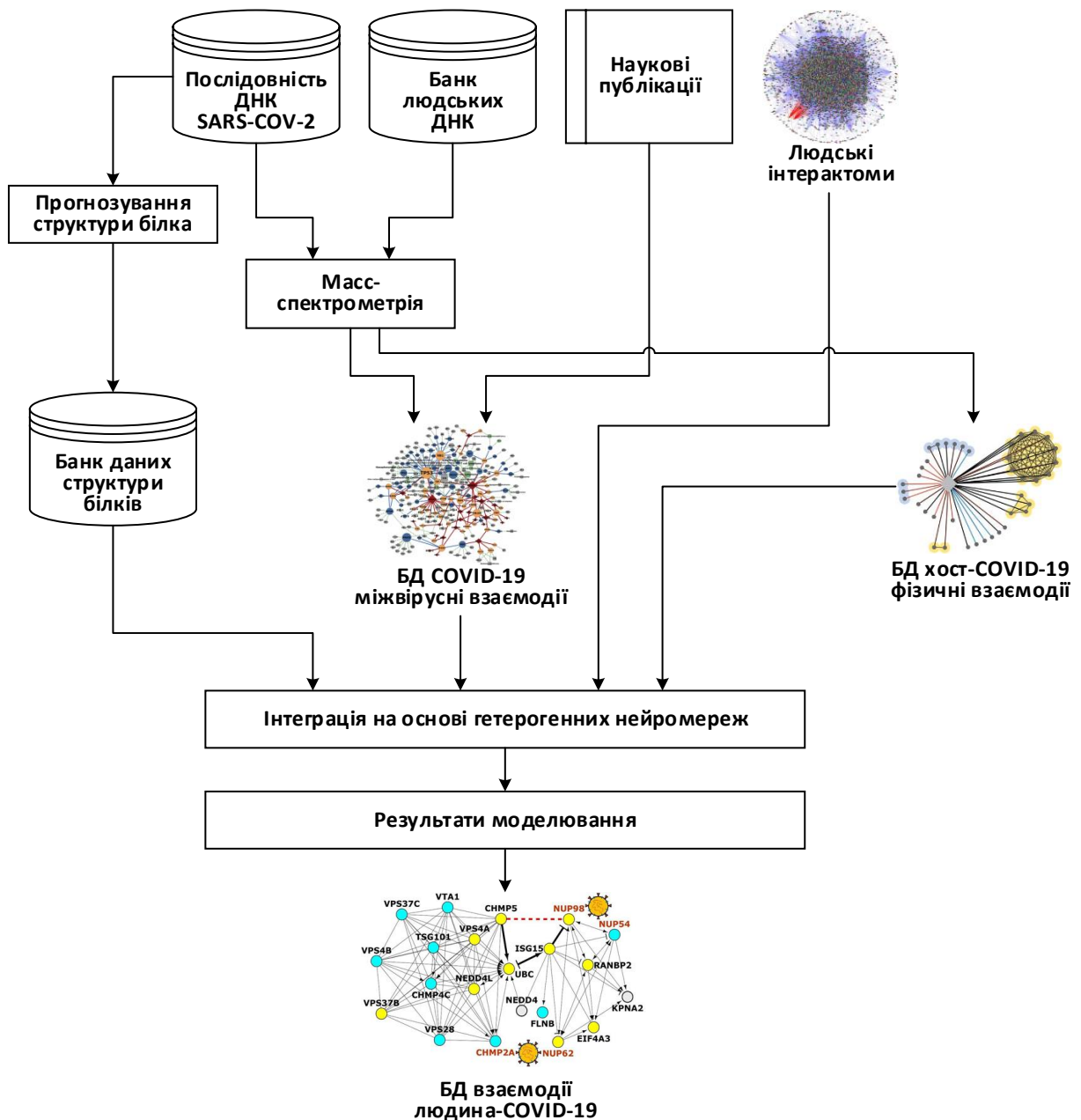


Рисунок 2.2 – Процес інтеграції даних для формування графічного подання інтерактом COVID-19 [48]

2.6 Аналіз зображень грудної клітки для діагностики та моніторингу COVID-19

Проведення аналітичного опрацювання зображень дозволяє перетворити цифрові зображення на дані, видобувши при цьому значущу інформацію щодо безпосередньо зображень. Автоматизоване аналітичне

опрацювання зображень грудної клітки допоможе при ранній діагностиці COVID-19. Внаслідок медичних спостережень було становлено, що для діагностики пневмонії у пацієнтів із COVID-19 є важливими два види зображень рентгенографії грудної клітки, отримані за допомогою рентгенографії та КТ-сканерів.

Візуалізація – це швидка, неінвазивна, відносно дешева та широко прийнята на даний час клінічна практика, яка може бути ефективно використана для моніторингу розвитку COVID-19. Її кінцева мета – це загальне покращення в галузі охорони здоров'я пацієнтів, дизайн біомаркерів COVID-19 та рання діагностика COVID-19 у громадян.

Для цього ефективно використовуються методи машинного навчання або методи, засновані на CNN. Зокрема, «COVID-Net» [50] є першим програмно-алгоритмічним комплексом, сформованим на основі CNN із відкритим кодом. Його розроблено з використанням методів глибинного навчання для підвищення ефективності процесів діагностування COVID-19.

У [51] подано опис методу легеневої КТ, котрий сформовано на основі аналітичних методів глибинного навчання, для виявлення випадків захворювання на пов'язану із COVID-19 пневмонію. Комбінована архітектура на основі довгої короткотривалої пам'яті (LSTM) та CNN використовується для аналітичного опрацювання рентгенівських зображень інфікованих COVID-19 пацієнтів [52].

Важливим джерелом інформації є спостереження за новими випадками COVID-19 котре використовується для прогнозування його еволюції на основі методів дифузії. При цьому використовувались дані з реального життя та сформовані на основі спеціальних математичних моделей [53].

Досконаліші моделі намагаються інтегрувати епідеміологічні дані щодо COVID-19 з іншими даними для дослідження впливу інших відомостей, зокрема, екологічних, географічних, клінічних [54]. В [55] подана модель прогнозування спалаху COVID-19, котра розроблена на основі LSTM-

мереж. Пандемія COVID-19 характеризується накопиченням значних обсягів епідеміологічних даних, котрі загальнодоступно поширюються в мережі в Інтернеті. Це, здебільшого, дані щодо інфікованих COVID-19 пацієнтів та летальних випадків (див. додаток А). При цьому слід відзначити, що набори таких даних мають ряд суттєвих недоліків:

- Низький рівень достовірності та надійності.
- Відсутність стандартизації форматів відомостей щодо COVID-19.
- Не стандартизовані семантичні формати даних.

На даний час не існує стандартизованого способу визначення точної кількості інфікованих COVID-19 громадян, але є критерії для класифікації ступенів тяжкості перебігу COVID-19 подані в додатку Б.

2.7 Аналітичне опрацювання в процесах використання та перепризначення лікарських препаратів для COVID-19

Аналіз даних з метою для виявлення можливих кандидатів серед існуючих лікарських засобів – це загально-відомий процес «перепризначення ліків». Він передбачає виявлення нових способів використання вже схвалених або експериментальних препаратів для можливого лікування нових патологій COVID-19.

Структуру процесу перепризначення медичних препаратів подано на рисунку 2.3. Процес перепризначення медичних препаратів базується на інтеграції молекулярних даних, зокрема, БД взаємодій між людиною і вірусом та БД взаємодії носіїв, що застосовуються поєднання лікарських препаратів та хвороб.

Наявність відомостей про цільові білки, взаємодії та асоціації між медичними препаратами та цільовими білками надає можливості швидкого пошуку відповідних молекул в БД препаратів, затверджених FDA, за допомогою вдосконалених моделей на основі глибинних нейронних мереж.

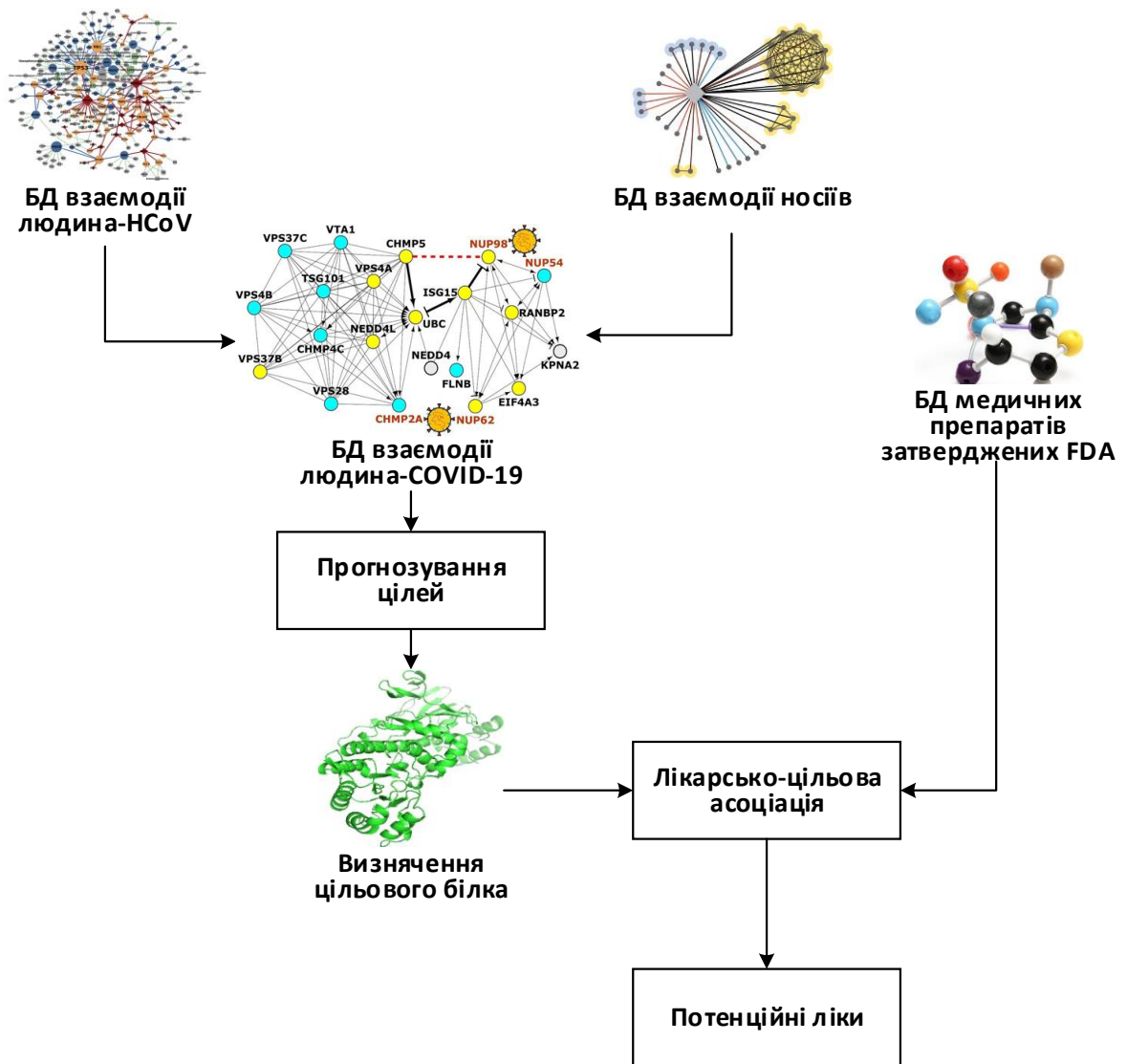


Рисунок 2.3 – Процес перепризначення лікарських препаратів

На даний час для значного числа вірусних захворювань доступний обширний перелік асоційованих цільових медичних препаратів та сформованих на основі методів глибинного навчання інструментів для перепризначення. Вони можуть ефективно використовуватись для аналітичного опрацювання відомостей щодо COVID-19 (див. таблицю 2.1). Останні тенденції досліджень спрямовані на дослідження методів вбудовування мереж [56] та DNN для складання списків можливих препаратів-кандидатів, які будуть підтвержені в лабораторних експериментах та клінічних дослідженнях.

Таблиця 2.1 – Методи аналітичного опрацювання відомостей щодо COVID-19 [6]

Задача	Тип даних	Моделі для аналітичного опрацювання даних
Філогенез/ вирівнювання.	Нуклеотидна/ білкова послідовність.	UPGMA, WPGMA, метод найближчих сусідів, максимальна ймовірність, метод Фітча-Марголіаша, максимальна парсимонія, байєсівські висновки.
Прогнозування структури.	Білкова послідовність.	Глибинні нейронні мережі (NeBcon, ResPRE, ResTriplet і TripletRes), QSQE, контрольоване машинне навчання (SVM), множинна регресія.
Прогноз SARS- CoV-2.	Нуклеотидна послідовність.	Звичайні моделі (наївний Байєс, K-найближчих сусідів, штучні нейронні мережі, дерева рішень, векторна підтримка), глибинні моделі CNN, Bi-path CNN (BiPathCNN).
Білкові взаємодії.	Білкова послідовність, мережі ІПП, структура білка.	Графічний аналіз.
Аналіз зображень грудної клітки.	Рентгенологічне або КТ- зображення грудної клітки.	Моделі глибинного навчання (VGG19), Mobile Net, Inception, Xception та Inception ResNet (v2,18,23,50), GAN, коефіцієнт схожості кісток (DSC).
Аналіз епідемічних тенденцій.	Експерименталь ний та спостережливий.	Статистичні моделі LSTM (SIR, байєсівська імпутація, лінійна та поліноміальна регресія).
Взаємодія з лікарськими засобами та перепризначення.	Білкова послідовність, молекули ліків, структура білка.	Графічний аналіз, графічні згорткові мережі.

Слід зазначити, що після ідентифікації «in silico» процес перепризначення препарату вимагає часу та фінансових затрат для проведення клінічних досліджень, але загальний час буде меншим чим при розробці нової молекули [57].

2.8 Аналітичне опрацювання метаболічних змін для важких та фатальних випадків COVID-19

Кумар Дас [58] опублікував відомості щодо дослідження наборів даних (див. додаток В) щодо клінічного стану пацієнтів з важким перебігом COVID-19, до яких включено відомості щодо летальних випадків. Основною метою дослідження було виявлення порушень метаболічних ланцюжків у чотирьох наборах даних. При цьому було встановлено п'ять відповідностей для найбільш змінених метаболічних ланцюжків для пацієнтів з важким та або летальним перебігом COVID-19 (див. рисунок 2.4).

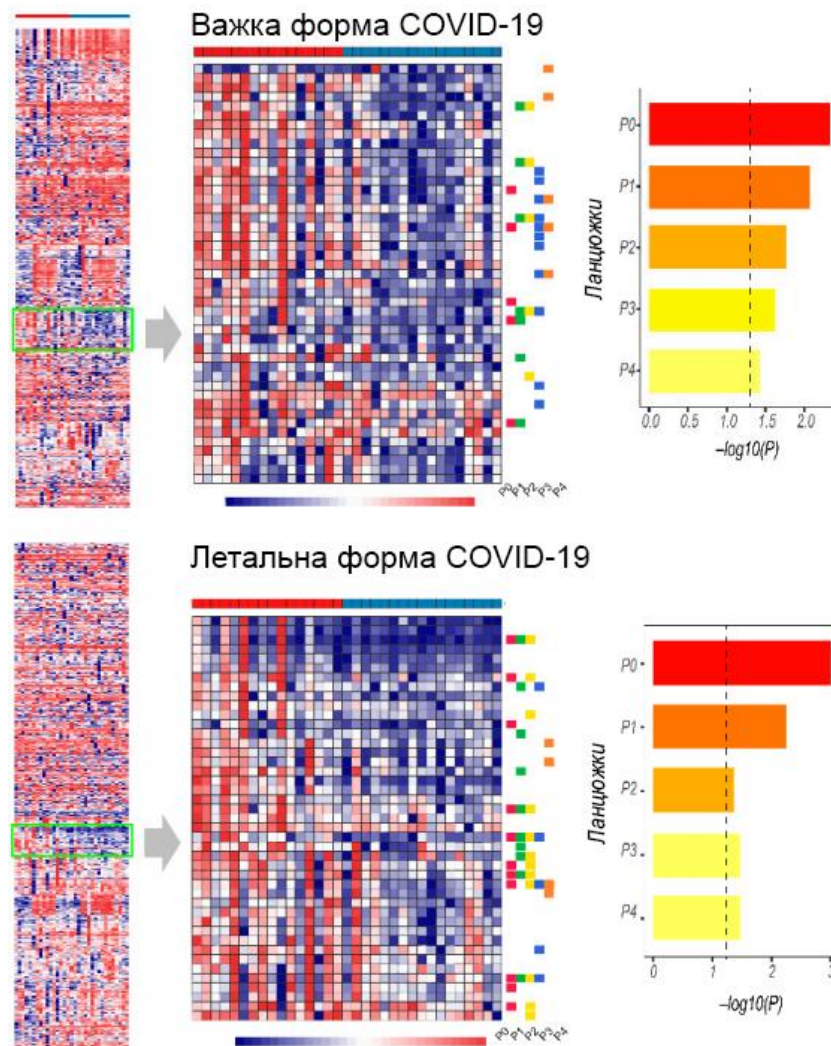


Рисунок 2.4 – Кластерний аналіз відомостей щодо груп пацієнтів з важким або летальним перебігом COVID-19 [58]

Автори [58] також провели аналітичне опрацювання трьох наборів даних з відомостями щодо летальних випадків COVID-19. Один з наборів даних було виключено, оскільки він містив лише два випадки та, як наслідок, був непридатним для проведення статистично-значимого аналізу. Тому метаболічні відмінності між важкими та летальними випадками COVID-19 (див. додаток Г) оцінювали за двома іншими наборами даних (див. рисунок 2.5). Із двох зазначених наборів даних було видобуто декілька спільних змін метаболічних ланцюжків. На основі яких за кластерною тепловою картою відбулося об'єднання двох значимих областей та, як наслідок, виділено три важливі метаболічні ланцюжки.

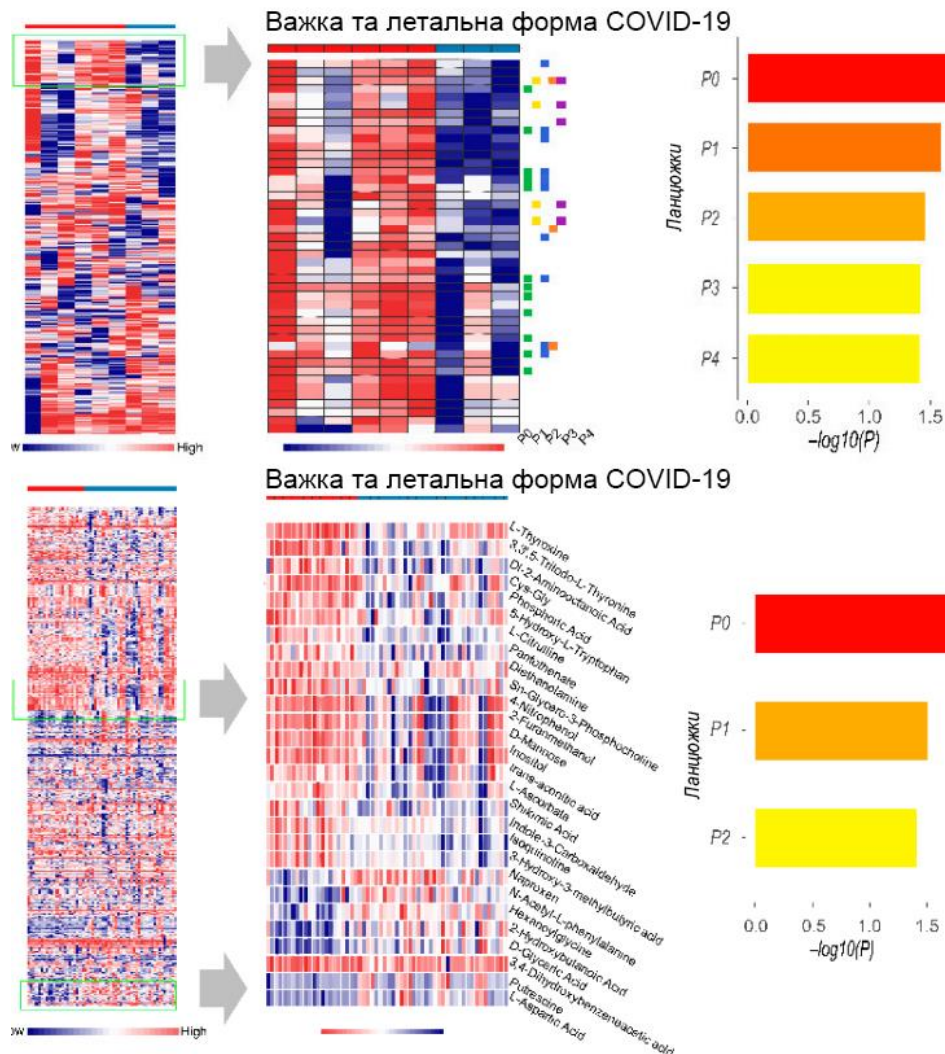


Рисунок 2.5 – Порівняльний кластерний аналіз відомостей між важкими та летальними групами пацієнтів COVID-19 [58]

COVID-19 може спричиняти патогенні зміни в дихальній системі людини, призвести до системних пошкоджень багатьох внутрішніх органів та смерті [59]. Запобігання летальних результатів COVID-19 є найважливішою метою сучасної клінічної практики та медичної допомоги. Окрім аналітичного опрацювання суттєвих порушень в обміні амінокислот вчені проводять дослідження для виявлення інших метаболічних порушень, спричинених COVID-19 [60], відслідковуючи при цьому метаболічні зміни при фатальних випадках COVID-19 [61]. Результати Апробації аналітичних моделей на основі кластерного аналізу (див. додаток Д) між COVID-19 та здоровими громадянами для семи наборів даних подана на рисунку 2.6. Після чого доцільно проводити аналіз ланцюжків на основі виявлених закономірностей.

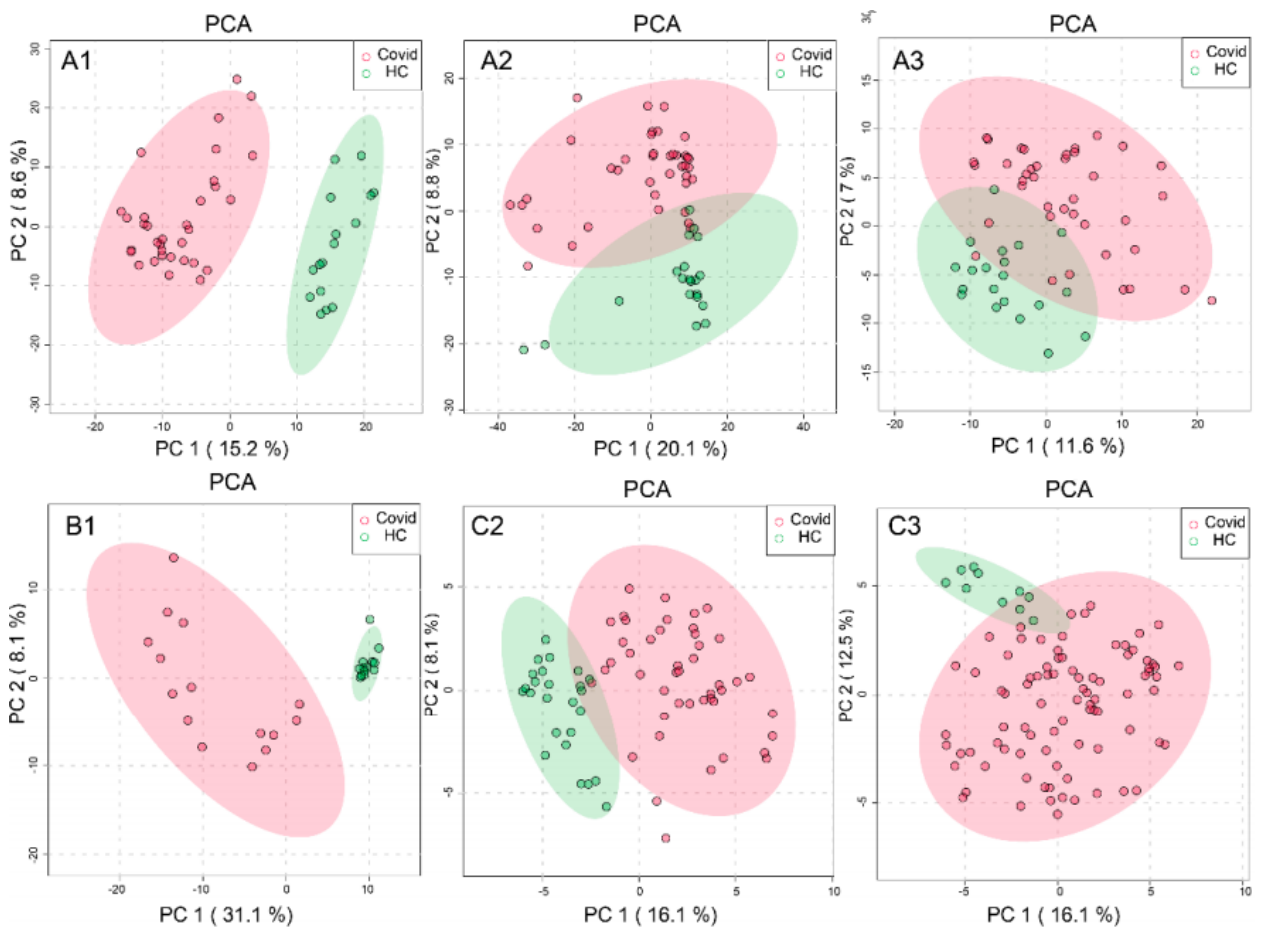


Рисунок 2.6 – Апробація аналітичних моделей на основі кластерного аналізу [58]

При цьому вдалося ідентифікувати кластери з послідовно-регульованими або знижено-регульованими метаболічними моделями між двома умовами для шести наборів даних.

Для глобальних наборів метаболомічних даних щодо COVID-19 характерні високі показники неоднорідності, що утруднює проведення мета-аналізу на рівні ознак. При цьому формуватиметься потенційне упередження, спричинене різницею у процедурах видобування та особливостях функціонування аналітичних платформ. Через можливість подання передбачуваних анотацій до відомостей щодо COVID-19, виявлені на даний час метаболічні ланцюжки потребують подальшої перевірки за допомогою системніших підходів сформованих з використанням множини аналітичних методів та засобів. На даний час малодосліджено потенційні значимі фактори, зокрема дієта та раціон харчування, етнічна приналежність, лікування тощо, оскільки більшість метаданих є неповними або відсутні в оригінальних дослідженнях. Для виявлення унікальних метаболічних ознак COVID-19 потрібно провести комплексні дослідження множини вірусних інфекцій COVID-19, SARS-CoV та грип [62].

З метою отримання загального огляд змін в метаболічній діяльності, спричинених COVID-19, автори [58] об'єднали всі, визначені на основі передбачуваних анотацій піків, значущі метаболіти у глобальну метаболічну мережу KEGG (див. рисунок 2.7) у якій коричневі вузли відображають відповідні метаболіти, розміри яких визначено на основі загальної кількості значень з різних наборів даних. Різнокольорові області представляють різні шляхи. Методи аналітичного опрацювання, сформовані на основі мережевих візуалізацій, можуть виявляти скоординовану метаболічну діяльність на основі скупчень метаболітів котрі розподілено у межах та поза межами ланцюжків. Внизу знаходяться п'ять видобутих ланцюжків, що відповідають п'яти кольоровим регіонам на рис. 2.7.

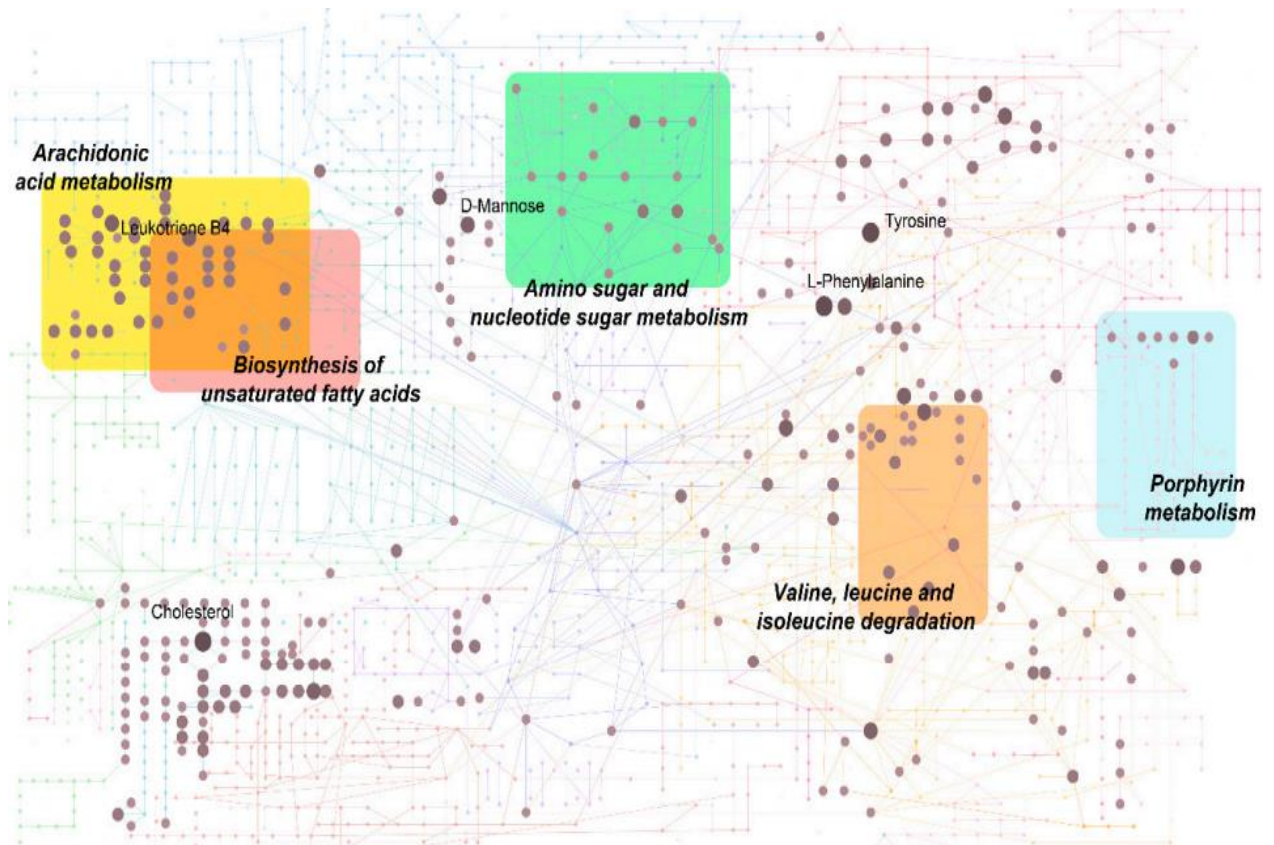


Рисунок 2.7 – Глобальна метаболічна мережа KEGG [58]

П'ять найкращих шляхів, було класифіковано за їх інтегрованими значеннями, їм відповідають п'ять кольорових областей на рис. 2.7 окреслюють п'ять найкращих шляхів, поданих на рисунку 2.8.

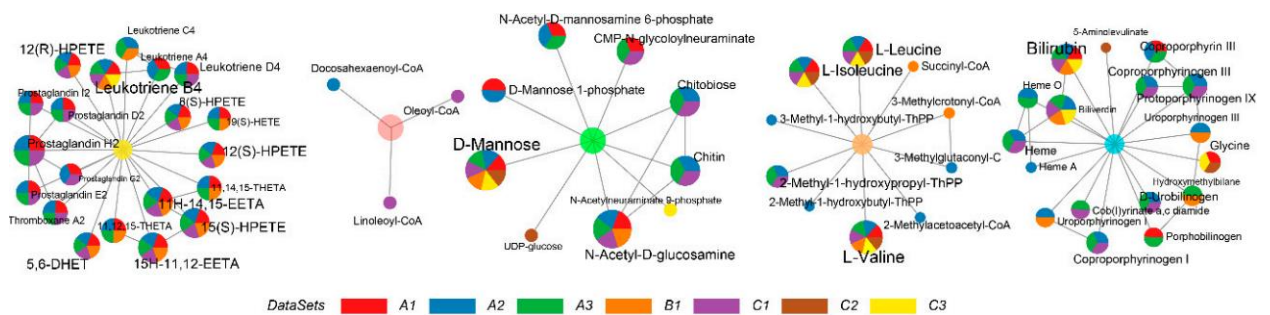


Рисунок 2.8 – Мережі потенційно порушених метаболічних ланцюжків [58]

При цьому вузли подано у формі кругових діаграм, розміри та компоненти яких відповідають включенням з різних наборів даних.

2.9 Висновок до другого розділу

В другому розділі кваліфікаційної роботи розглянуто метааналіз глобальних наборів даних щодо COVID-19. Описано аналітичне опрацювання даних SARS-CoV-2 на основі відомостей БД Omics. Розглянуто філогенний та мутаційний варіаційний аналіз даних щодо COVID-19. Досліджено аналітичне опрацювання генетичної інформації щодо COVID-19. Подано особливості інтерактивного аналізу відомостей щодо COVID-19. Висвітлено аналіз зображень грудної клітки для діагностики та моніторингу COVID-19. Описано аналітичне опрацювання в процесах використання та перепризначення лікарських препаратів для COVID-19. Зокрема, в параграфі наведено класифікаційну таблицю методів аналітичного опрацювання відомостей щодо COVID-19. Подано описи досліджень щодо аналітичного опрацювання метаболічних змін для важких та фатальних випадків COVID-19.

3 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ХОРОНИ ПРАЦІ

3.1 Актуальність безпеки життєдіяльності людини

Поширення кризових явищ внаслідок пандемії COVID-19 ускладнює умови професійної діяльності працівників практично всіх галузей та підвищує актуальність безпеки життєдіяльності людини.

Нормативно правова база в галузі безпеки життєдіяльності людини (відповідно до Конституції України) забезпечується законами України: «Про основи національної безпеки» від 19 червня 2003 р. за № 964; «Про захист населення і територій від надзвичайних ситуацій техногенного та природного характеру» від 08 червня 2000 р. за № 1809; «Про внесення змін до Закону України «Про охорону праці» від 21 грудня 2002 р. за № 229; «Про правовий режим надзвичайного стану» від 16 березня 2000 р. за № 1550; Концепція освіти з напрямку «Безпека життя і діяльності людини» від 12 березня 2001 р.; а також навчальними програмами нормативної дисципліни «Безпека життєдіяльності» для студентів вищих навчальних закладів відповідних освітніх рівнів усіх спеціальностей [63].

За даними статистичних результатів 73% нещасних випадків та аварій на виробництві відбулися внаслідок організаційних причин, при цьому лише 14% – з технічних та 13% – з психофізіологічних. Отже умови сучасного життя вимагають чіткого розуміння людиною, як правильно поводитися в умовах будь-якої надзвичайної ситуації.

Безпека життєдіяльності (БЖД) – це наука, яка вивчає проблеми безпечного перебування людини в довкіллі у процесі здійснення різних видів її діяльності, зокрема трудової. Зазначена галузь є універсальною дисципліною тому що «Охорона праці» та «Цивільна оборона» розглядають лише окремі випадки безпеки у визначених ситуаціях. Охорона праці досліджує людину в умовах виробництва а цивільна оборона – в

надзвичайних ситуаціях. Безпека життєдіяльності ж досліджує людину в усіх життєвих обставинах та ситуаціях [64].

Вся множина процесів людської активності формує поняття діяльності, котра вирізняє людину від решти живих істот. Людська діяльність є специфічною формою активності, котра необхідна для людського суспільства. На даний час доступні найрізноманітніші форми людської діяльності, зокрема практичні, інтелектуальні та духовні процеси в побуті, громадській, культурній, виробничій, науковій та інших сферах життя та суспільства. Навчальний курс «Безпека життєдіяльності» призначений:

- Для сприяння усвідомлення того, що на першому місці повинна бути людина, як головна суспільна цінність, виховання гуманного та усвідомленого відношення до питань особистої безпеки та навколишньої безпеки в усіх сферах людської діяльності та відносин.

- Ознайомлення з інноваціями щодо проектування нової техніки та технологій відповідно до сучасних екологічних вимог з урахуванням вимог щодо стійкості функціонування господарських об'єктів та технічних систем;

- Прогнозувати можливі зміни в умовах надзвичайних ситуацій та приймати обґрунтовані та ефективні рішення щодо захисту населення та персоналу об'єктів від можливих негативних наслідків.

- Забезпечити високоефективне формування стереотипу поведінки людини з метою виживання в природних та антропогенних умовах.

Метою вивчення курсу «Безпека життєдіяльності» у вищих навчальних закладах є формування основ культури та поведінки людини щодо безпеки, формування відповідних моральних цінностей, поглядів та поведінки; забезпечення належного стану особистої захищеності завдяки формуванню необхідної множини знань, вмінь та навичок щодо реагування на надзвичайні ситуації природного, техногенного, екологічного та соціального типів; формування правильної світоглядної та соціальної позиції щодо умов власної безпеки, мотивація для безпечної поведінки в побуті, на виробництві тощо.

3.2 Долікарська допомога при пораненнях

При будь-якому порушенні цілісності шкіри та глибоких враженнях тканин організму людини необхідно обробити шкіру навколо рани розчином йоду, спиртом тощо. При цьому не рекомендується промивати рану водою або дезінфікуючими розчинами. Після завершення процесу оброблення рани слід накласти асептичну пов'язку. Пов'язка призначена для захисту рани від можливого забруднення, інфікування, зменшуючи при цьому біль. Зовнішній вигляд перев'язаної рани заспокоює хворого.

В процесі обробки рани потрібно дотримуватись наступних правил:

- Перед початком обробки рани необхідно ретельно помити руки. Якщо поряд немає води, то їх слід протерти спиртом або бензином.
- Невеликі поранення або садна, після оброблення шкірного покриву навколо них настоячкою йоду або перекисом водню, заклеюють лейкопластиром або біомедичним клеєм БФ-6.
- З ран не можна видаляти сторонні тіла або бруд, оскільки при цьому можна пошкодити судини та викликати сильну кровотечу.
- Шкірний покрив навколо рани слід протирати у напрямку від країв до периферії за допомогою шматочком марлі, бинта або ватним тампоном, котрі мають бути змочені спиртом, спиртовим розчином йоду або бензином. При цьому не можна заливати рану йодом.
- Слід зробити салфетку з використанням бинта або індивідуального пакета. При цьому розміри салфетки мають бути такими, щоб вона повністю закривала всю рану. Салфетку слід накласти поверхню рани та забинтувати або приклеїти смужками лейкопластиру.
- Якщо в рані візуально спостерігаються внутрішні органи, мозок або сухожилля, то потрібно акуратно накласти стерильну пов'язку, щоб у рану часом не потрапила інфекція. Краще накрити рану стерильним матеріалом.

При проникаючих пораненнях тіла людини перша допомога спрямована на запобігання можливості інфікування рани. У випадках проникаючих поранень грудної клітки може відбуватись кровохаркання, проникнення повітря під шкіру, яке виглядає наче набряк, але хрустить при натисненні, накопичення повітря у плевральній порожнині (пневмоторакс), накопиченням крові у плевральній порожнині (гемоторакс).

Пневмоторакс – це накопичення атмосферного повітря в плевральній порожнині, яке потрапляє через відкрите поранення грудної клітки. В такому випадку – це відкритий пневмоторакс, а при ушкодженні легені чи бронха – закритий пневмоторакс. При його виникненні можливе балотування органів середостіння, що супроводжується розладами циркуляції крові та дихання. При проникаючих пораненнях грудної клітки перша допомога повинна бути спрямована на усунення пневмотораксу, попередження травматичного шоку та захист рани від попадання інфекцій. Здебільшого проникаючі поранення органів черевної порожнини є внаслідками вогнепальних та колото-різаних ран. Для всіх поранень черевної порожнини характерними ознаками є різкий біль у животі, напруження м'язів черевної живота, симптоми внутрішньої кровотечі, шоку та колапсу. Об'ємна кровотеча супроводжує ушкодження паренхіматозних органів, зокрема печінки, селезінки, нирок тощо. У всіх випадках проникаючих поранень черевної порожнини необхідно обробити рану і накласти асептичну пов'язку. Прийом води і їжі категорично забороняється. Пораненого у живіт необхідно негайно госпіталізувати, транспортувати його у лежачому положенні.

3.3 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи розглянуто актуальність безпеки життєдіяльності людини та описана долікарська допомога при пораненнях.

ВИСНОВКИ

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр»:

– Проведено аналіз предметної області.

– Описано джерела даних щодо COVID-19 та подано їх інтегральну характеристику.

– Висвітлено конвеєрне опрацювання даних щодо COVID-19.

В другому розділі кваліфікаційної роботи:

– Розглянуто метааналіз глобальних наборів даних щодо COVID-19.

– Описано аналітичне опрацювання даних SARS-CoV-2 на основі відомостей БД Omics.

– Розглянуто філогенний та мутаційний варіаційний аналіз даних щодо COVID-19.

– Досліджено аналітичне опрацювання генетичної інформації щодо COVID-19.

– Подано особливості інтерактивного аналізу відомостей щодо COVID-19.

– Висвітлено аналіз зображень грудної клітки для діагностики та моніторингу COVID-19.

– Описано аналітичне опрацювання в процесах використання та перепризначення лікарських препаратів для COVID-19. Зокрема, наведено класифікаційну таблицю методів аналітичного опрацювання відомостей щодо COVID-19.

– Подано описи досліджень щодо аналітичного опрацювання метаболічних змін для важких та фатальних випадків COVID-19.

У розділі «Безпека життєдіяльності, основи хорони праці» висвітлено актуальність безпеки життєдіяльності людини та описано долікарську допомогу при пораненнях.

ПЕРЕЛІК ДЖЕРЕЛ

- 1 Duda O., Kunanets N., Matsiuk O., Pasichnyk V., Rzhenskyi A. (2021) Aggregation, Storing, Multidimensional Representation and Processing of COVID-19 Data. In: Shakhovska N., Medykovsky M.O. (eds) *Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing*, vol 1293, pp 875-889. Springer, Cham. ISBN978-3-030-63270-0.
- 2 Duda, O., Pasichnyk, V., Kunanets, N., Antonii, R., & Matsiuk, O. (2020, September). Multidimensional Representation of COVID-19 Data Using OLAP Information Technology. In *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 2, pp. 277-280)*. IEEE.
- 3 Duda, O., Palka, O., Pasichnyk, V., Matsiuk, O., Kunanets, N., & Tabachyshyn, D. (2020, September). Existing City Assessment Systems. In *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 2, pp. 238-241)*. IEEE.
- 4 Johns Hopkins. Coronavirus resource center. <https://coronavirus.jhu.edu/map.html>.
- 5 Duda, O., et al, Selection of Effective Methods of Big Data Analytical Processing in Information Systems of Smart Cities. *CEUR Workshop Proceedings 2631*, pp. 68-78. 2020.
- 6 Kumar Das, Jayanta, et al. "Data science in unveiling COVID-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing." *Briefings in Bioinformatics* 22.2 (2021): 855-872.
- 7 Duda, O., Kunanets, N., Martsenko, S., Matsiuk, O., Pasichnyk, V., Building secure Urban information systems based on IoT technologies. *CEUR Workshop Proceedings 2623*, pp. 317-328. 2020.

- 8 Noruzi A. Google scholar: the new generation of citation indexes. *Libri* 2005;55(4):170–80.
- 9 Dimensions. <https://app.dimensions.ai/discover/publication>.
- 10 Omics. <http://omics.org>.
- 11 Veltri P, Cannataro M, Tradigo G. Sharing mass spectrometry data in a grid-based distributed proteomics laboratory. *Inf Process Manag* 2007;43(3):577–91. cited By 19.
- 12 NCBI. <http://www.ncbi.nlm.nih.gov/geo/>.
- 13 OmicsDI. <https://www.omicsdi.org/>.
- 14 Srinivasan S, Cui H, Gao Z, *et al.* Structural genomics of sarscov-2 indicates evolutionary conserved functional regions of viral proteins. *Viruses* 2020;12(4):360.
- 15 Gordon DE, Jang GM, Bouhaddou M, *et al.* A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;583(7816):459–68.
- 16 Perfetto L, Pastrello C, Del-Toro N, *et al.* The imex coronavirus interactome: an evolving map of coronaviridae–host molecular interactions. *BioRxiv* 2020.
- 17 Zhou S, Wang Y, Zhu T, *et al.* Ct features of coronavirus disease 2019 (covid-19) pneumonia in 62 patients in Wuhan, China. *Am J Roentgenol* 2020;214(6):1287–94.
- 18 TrainingData.io. <https://www.trainingdata.io/>.
- 19 Dong E, Hongru D, Gardner L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect Dis* 2020;20(5):533–4.
- 20 Italian Data. <https://github.com/pcm-dpc/COVID-19>.
- 21 Xu B, Moritz UG, Kraemer BG, *et al.* Open access epidemiological data from the covid-19 outbreak. *Lancet Infect Dis* 2020;20(5):534.
- 22 Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* 2006;5(10):821–34.

- 23 Réda C, Kaufmann E, Delahaye-Duriez A. Machine learning applications in drug development. *Comput Struct Biotechnol J* 2020;18:241–52.
- 24 Voytek B. Social media, open science, and data science are inextricably linked. *Neuron* 2017;96(6):1219–22.
- 25 Roy S, Sharma P, Nath K, *et al.* Pre-processing: a data preparation step. *Encyclop Bioinform Comput Biol ABC Bioinform* 2018;463.
- 26 Jha M, Guzzi PH, Veltri P, *et al.* Functional module extraction by ensembling the ensembles of selective module detectors. *Int J Comput Biol Drug Design* 2019;12(4):345–61.
- 27 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
- 28 Kipf, TN, Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- 29 Waheed A, Goyal M, Gupta D, *et al.* Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access* 2020;8:91916–23.
- 30 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. *Adv Neural Inf Process Syst* 2014;2672–80.
- 31 Ghosal S, Sengupta S, Majumder M, *et al.* Linear regression analysis to predict the number of deaths in India due to sars-cov-2 at 6 weeks from day 0 (100 cases–March 14th 2020). *Diab Metab Syndrome Clin Res Rev* 2020;14(4):311–5.
- 32 Roy S, Bhattacharyya DK. An approach to find embedded clusters using density based techniques. In: *International Conference on Distributed Computing and Internet Technology*. Springer, 2005, 523–35.
- 33 Zhou, G.; Stevenson, M.M.; Geary, T.G.; Xia, J. Comprehensive Transcriptome Meta-analysis to Characterize Host Immune Responses in Helminth Infections. *PLoS Negl. Trop. Dis.* 2016, 10, e0004624.

34 Gardinassi, L.G.; Souza, C.O.S.; Sales-Campos, H.; Fonseca, S.G. Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. *Front. Immunol.* 2020, *11*, 1636.

35 Ceraolo C, Giorgi FM. Genomic variance of the 2019-ncov coronavirus. *J Med Virol* 2020;92(5):522–8.

36 Gonzalez JM, Gomez-Puertas P, Cavanagh D, *et al.* A comparative sequence analysis to revise the current taxonomy of the family coronaviridae. *Arch Virol* 2003;148(11):2207–35.

37 Yavarian J, Rezaei F, Shadab A, *et al.* Cluster of Middle East respiratory syndrome coronavirus infections in Iran, 2014. *Emerg Infect Dis* 2015;21(2):362.

38 Linnon Biosoft. Introduction to DNAMAN Sequence Analysis Functions. <https://www.lynnon.com/dnaman.html>.

39 Wu C, Yang L, Yang Y, *et al.* Analysis of therapeutic targets for sars-cov-2 and discovery of potential drugs by computational methods. *Acta Pharm Sinica B* 2020;10(5):766–88.

40 Zhang T, Wu Q, Zhang Z. Probable pangolin origin of sars-cov-2 associated with the covid-19 outbreak. *Curr Biol* 2020;30(7):1346–51.

41 Waterhouse AM, Procter JB, Martin DMA, *et al.* Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25(9):1189–91.

42 Katoh K, Standley DM. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772–80.

43 Turista DDR, Islamy A, Kharisma VD, *et al.* Distribution of covid-19 and phylogenetic tree construction of sars-cov-2 in Indonesia. *J Pure Appl Microbiol* 2020;14(suppl 1):1035–42.

44 Yin C. Genotyping coronavirus sars-cov-2: methods and implications. *Genomics* 2020;112(5):3588–96.

45 Fang Z, Tan J, Wu S, *et al.* Ppr-meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience* 2019;8(6) giz066.

46 Chi X, Liu X, Wang C, *et al.* Humanized single domain antibodies neutralize sars-cov-2 by targeting the spike receptor binding domain. *Nat Commun* 2020;11(1):1–7.

47 Perfetto L, Pastrello C, Del-Toro N, *et al.* The imex coronavirus interactome: an evolving map of coronaviridae–host molecular interactions. *BioRxiv* 2020.

48 Messina F, Giombini E, Agrati C, *et al.* Covid-19: viral–host interactome analyzed by network based-approach model to study pathogenesis of sars-cov-2 infection. *J Transl Med* 2020;18(1):1–10.

49 Das J, Chakrobarty S, Roy S. Impact analysis of sarscov2 on signaling pathways during covid19 pathogenesis using codon usage assisted host–viral protein interactions. *bioRxiv* 2020.

50 Kim M, Kang J, Kim D, *et al.* Hi-covidnet: deep learning approach to predict inbound covid-19 patients and case study in South Korea. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, 3466–73.

51 Butt C, Gill J, Chun D, *et al.* Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl Intell* 2020;1.

52 Islam MZ, Islam MM, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (covid-19) using x-ray images. *Inform Med Unlock* 2020;100412.

53 Giordano G, Blanchini F, Bruno R, *et al.* Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 2020;1–6.

54 Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to covid-19 in Italy. *JAMA* 2020;323(18):1775–6.

55 Chimmula VKR, Zhang L. Time series forecasting of covid19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* 2020;109864.

56 Zhou Y, Hou Y, Shen J, *et al.* Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell Discov* 2020;6(1):1–18.

57 Harrison C. Coronavirus puts drug repurposing on the fast track. *Nat Biotechnol* 2020;38(4):379–81.

58 Pang, Zhiqiang, *et al.* "Comprehensive Meta-Analysis of COVID-19 Global Metabolomics Datasets." *Metabolites* 11.1 (2021): 44.

59 Bradley, B.T.; Maioli, H.; Johnston, R.; Chaudhry, I.; Fink, S.L.; Xu, H.; Najafian, B.; Deutsch, G.; Lacy, J.M.; Williams, T.; *et al.* Histopathology and ultrastructural findings of fatal COVID-19 infections in Washington State: A case series. *Lancet* 2020, 396, 320–332.

60 Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* 2020, 182, 59–72.e15.

61 Polonikov, A. Endogenous Deficiency of Glutathione as the Most Likely Cause of Serious Manifestations and Death in COVID-19 Patients. *ACS Infect. Dis.* 2020, 6, 1558–1562.

62 Thair, S.A.; He, Y.D.; Hasin-Brumshtein, Y.; Sakaram, S.; Pandya, R.; Toh, J.; Rawling, D.; Remmel, M.; Coyle, S.; Dalekos, G.N.; *et al.* Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections. *iScience* 2021, 24, 101947.

63 Пліско, Ю. В., and Ю. Пасечник. "Актуальність вивчення курсу Основи безпеки життєдіяльності та охорони праці у вищих навчальних закладах." *Духовність особистості: методологія, теорія і практика* 3 (2018): 181-188.

64 ГРИГА, Г.Ю. "АКТУАЛЬНІ ПРОБЛЕМИ БЕЗПЕКИ ЖИТТЄДІЯЛЬНОСТІ."

ДОДАТКИ

**Узагальнені клінічно-демографічні характеристики множини зразків
COVID-19 [55]**

Характеристики	Тяжкість перебігу			
	Здоровий	Легка та середня	Тяжка	Летальна
Вік, років				
<20	0	4	0	0
20~60	69	96	36	0
> 60	3	40	23	4
Не зрозуміло*	56	28	22	36
Стать				
Чоловіча	50	72	39	7
Жіноча	33	57	19	1
Не зрозуміло**	26	28	22	36
Час початку, дні				
Лікування				
Гідроксихлорохін	-	28	-	-
Ремдесивір	-	3	-	-

* Не зрозуміло: відповідна інформація недоступна.

** Не зрозуміло: «Кількість» стосується кількості зразків, а не кількості осіб.

Критерії для класифікації ступенів тяжкості перебігу COVID-19

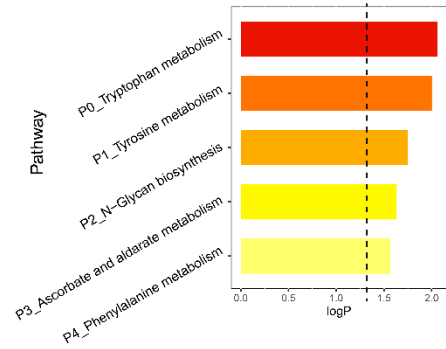
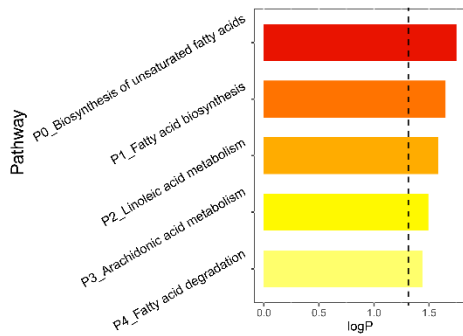
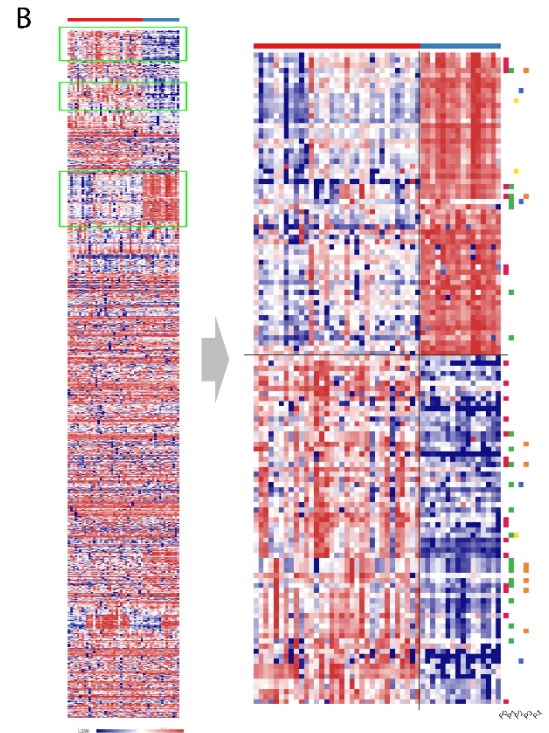
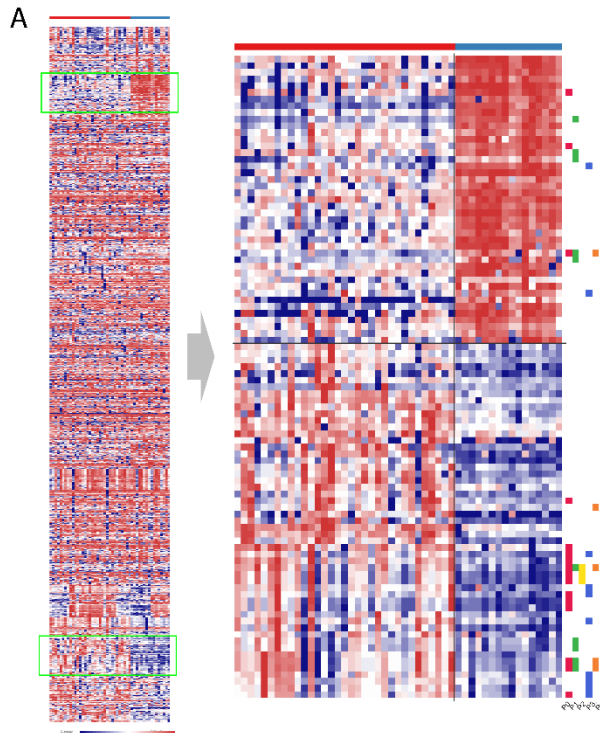
Важкість перебігу COVID-19	Критерії класифікації
Від легкого до помірного	Госпіталізований, але без кисню або з киснем із низьким рівнем потоку.
	Не виявлені імагологічні зміни, або виявлені, але такі що не відповідають стандарту тяжкості.
	Концентрація ІЛ-6 у сироватці крові менше 90 пг/мл.
Важкий	Інтубація та вентиляція легень або великий потік кисню, або додаткова підтримка внутрішніх органів.
	Задишка, RR \geq 30 уд./хв.
	SO ₂ менше 93% у стані спокою.
	PaO ₂ /FiO ₂ менше 300 мм рт. ст.
	Гострий прогрес з імагологічним ураженням зріс більш ніж на 50% за останні 24-48 годин.
	Шок.
	Концентрація ІЛ-6 у сироватці крові більше 90 пг/мл.
Фатальний	Повідомлення про смерть як клінічний кінець.
Не COVID-19	Проявились подібні клінічні характеристики, включаючи лихоманку та/або кашель, як у пацієнтів з COVID-19, але негативні результати тестування на нуклеїнову кислоту.

Технічна інформація щодо наборів даних [58]

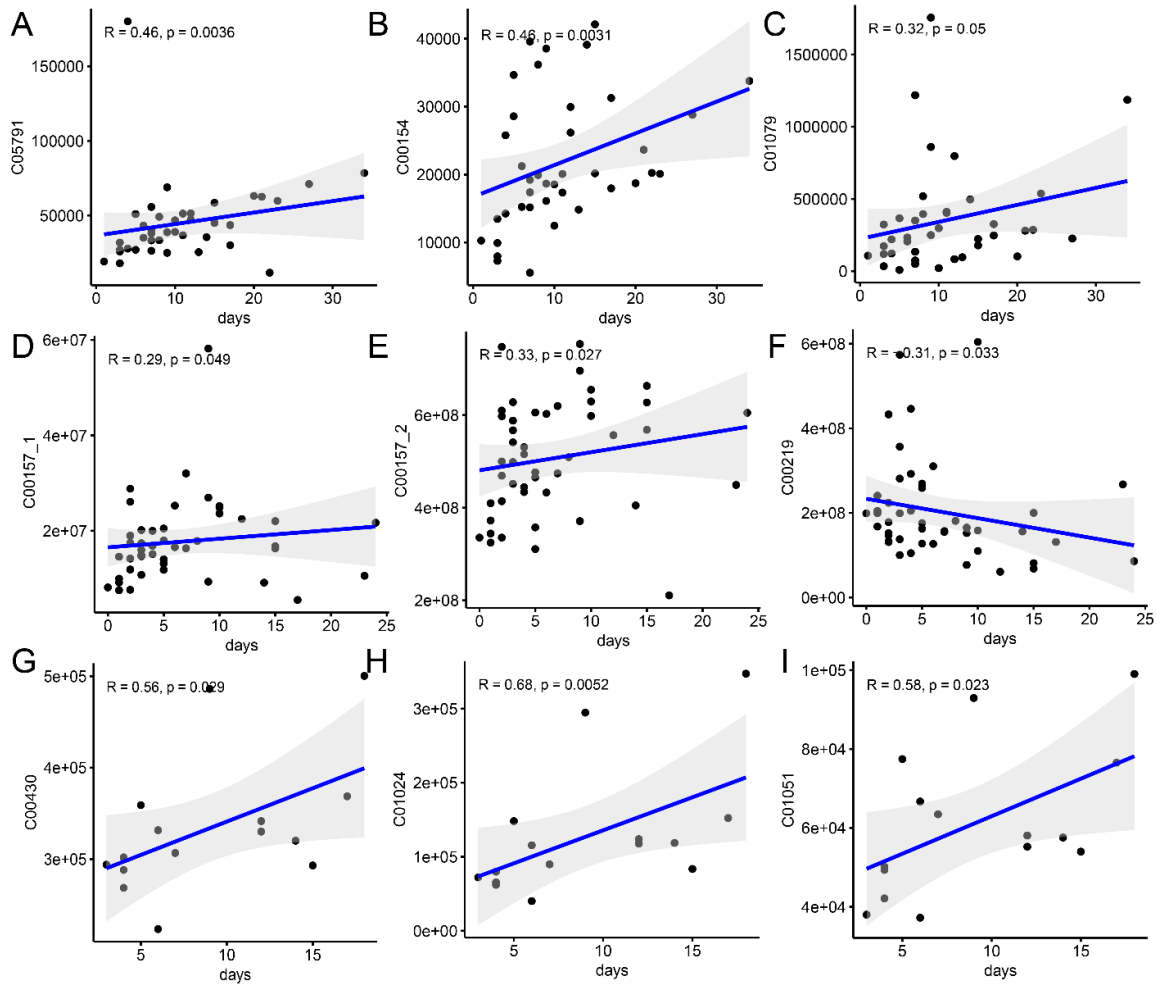
Код	Тип даних	Платформа	Стандарт	Основні видобуті відомості*	Іонний режим	Зразок
A1	Raw	UPLC-QE	C18	Фольк -> Хлороформ/метанол	Нег. + Поз.	Сироватка
A2	Raw	UPLC- QTOF	HILIC	Ацетон + метанол -> ацетонітрил + H ₂ O	Нег. + Поз.	Сироватка
A3			C18			
C1	Raw	UPLC-TTOF	C18	Хлороформ/метанол - > вода -> водний метанол	Нег. + Поз.	Плазма
C2	Table	UPLC- QEHF	C18	Етанол -> метанол	Нег. + Поз.	Сироватка
C3	Table	UPLC- QTRAP	C30	Метанол + МТВЕ	Нег. + Поз.	Плазма
B1	Raw	HPLC- micrOTOF	C18	MeOH:MeCN -> Вода	Нег. + Поз.	Плазма

* Метод видобування біологічного матеріалу, лише коротко показує основні етапи досліджень.

Кластерний аналіз теплових карт між групами Covid-19 та здорових осіб (НС) у негативному (А) та позитивному режимах (В) [58]



**Кореляційний аналіз спірмена щодо часу настання (днів) з метаболітами
в значно порушених ланцюжках [58]**



- A. C05791, D-уробіліноген.
- B. C00154, Пальмітоїл-КоА.
- C. C01079, протопорфіриноген ІХ.
- D. C00157, ПК (16: 0/16: 1 (9Z)).
- E. C00157, ПК (16: 0/18: 0).
- F. C00219, арахідонова кислота.
- G. C00430, 5-амінолевулінат.
- H. C01024, гідроксиметилбілан.
- I. C01051, Уропорфіриноген ІІІ.