

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

магістр

(назва освітнього ступеня)

на тему: Методи і засоби аналітичного опрацювання відомостей щодо COVID-19

Виконав: студент VI курсу, групи СНм-61
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)

(підпис)

Ваник А.Г.

(прізвище та ініціали)

Керівник

(підпис)

Мацюк О.В.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Мацюк О.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Гащин Н.Б.

(прізвище та ініціали)

Тернопіль
2020

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)

«23» грудня 2020 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

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

Студенту Ваник Анастасії Геннадіївни
(прізвище, ім'я, по батькові)

1. Тема роботи Методи і засоби аналітичного опрацювання відомостей щодо COVID-19

Керівник роботи Мацюк О.В., к.т.н., доцент кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «___» _____ 20__ року № _____

2. Термін подання студентом завершеної роботи 24 грудня 2020р.

3. Вихідні дані до роботи Наукові публікації про методи та засоби аналітичного опрацювання відомостей щодо COVID-19.

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ. 1 Актуальність аналітичного опрацювання відомостей щодо COVID-19. 2 Техніки видобування даних щодо COVID-19. 3 Практики використання аналітичних інструментів для опрацювання відомостей щодо COVID-19. 4 Охорона праці та безпека в надзвичайних ситуаціях. Висновки.

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

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Дмитроца Л.П., доцент		
Безпека в надзвичайних ситуаціях	Стадник І.П., професор		

7. Дата видачі завдання 21 вересня 2020 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	21.09.20-27.09.20	<i>Виконано</i>
2.	Підбір наукових джерел методи та засоби аналітичного опрацювання відомостей щодо COVID-19	28.09.20-04.10.20	<i>Виконано</i>
3.	Переклад та опрацювання наукових джерел про методи та засоби аналітичного опрацювання відомостей щодо COVID-19	05.10.20-11.10.20	<i>Виконано</i>
4.	Виконання дослідження щодо формування методів та застосування засобів аналітичного опрацювання відомостей щодо COVID-19	12.10.20-18.10.20	<i>Виконано</i>
5.	Оформлення розділу «Актуальність аналітичного опрацювання відомостей щодо COVID-19»	19.10.20-25.10.20	<i>Виконано</i>
6.	Оформлення розділу «Техніки видобування даних щодо COVID-19»	26.10.20-01.11.20	<i>Виконано</i>
7.	Оформлення розділу «Практики використання аналітичних Інструментів для опрацювання відомостей щодо COVID-19»	02.11.20-08.11.20	<i>Виконано</i>
8.	Виконання завдання до підрозділу «Охорона праці»	09.11.20-15.11.20	<i>Виконано</i>
9.	Виконання завдання до підрозділу «Безпека в надзвичайних ситуаціях»	16.11.20-22.11.20	<i>Виконано</i>
10.	Оформлення кваліфікаційної роботи	23.11.20-29.11.20	<i>Виконано</i>
11.	Нормоконтроль	30.11.20-05.12.20	<i>Виконано</i>
12.	Перевірка на плагіат	07.12.20	<i>Виконано</i>
13.	Попередній захист кваліфікаційної роботи	10.12.20	<i>Виконано</i>
14.	Захист кваліфікаційної роботи	24.12.20	

Студент

(підпис)

Ваник А.Г.

(прізвище та ініціали)

Керівник роботи

(підпис)

Мацюк О.В.

(прізвище та ініціали)

АНОТАЦІЯ

Методи і засоби аналітичного опрацювання відомостей щодо COVID-19 // Кваліфікаційна робота освітнього рівня "Магістр" // Ваник Анастасія Геннадіївна // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНм-61 // Тернопіль, 2020 // С. 80, рис. – 8, табл. – 2, кресл. – , додат. – 1, бібліогр. – 128.

Ключові слова: видобування, дані, інструмент, інтеграція, конфіденційність, опрацювання, охорона здоров'я.

Кваліфікаційна робота присвячена дослідженню методів та засобів аналітичного опрацювання відомостей щодо COVID-19. В першому зафіксовано стан предметної області. Проаналізовано видобування даних щодо COVID-19. Висвітлено етапи попереднього їх опрацювання. В другому розділі досліджено методи видобування відомостей щодо COVID-19. Розглянуто збирання та попереднє опрацювання соціальних відомостей щодо COVID-19. Проаналізовано особливості інтеграції даних щодо COVID-19. Описано інструменти для видобування відомостей щодо COVID-19. В третьому розділі кваліфікаційної роботи досліджено злиття та видобування відомостей щодо COVID-19 отриманих з ІоН-джерел. Описано використання штучного інтелекту для опрацювання відомостей щодо COVID-19. Проаналізовано методи навчання систем штучного інтелекту при діагностуванні COVID-19. У розділі «Охорона праці та безпека в надзвичайних ситуаціях» описано нормативні документи, які визначають дії роботодавця, працівник якого захворів на COVID-19 та підвищення стійкості роботи об'єктів господарської діяльності у воєнний час.

ANNOTATION

Methods and facilities of COVID-19 data analytical processing // qualification work of «Master» degree // Vanyk Anastasia Gennadyevna // Ternopil National Technical University named after Ivan Pulyuy, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science Sciences, group SNm-61 // Ternopil, 2020 // P. – 80, fig. – 8, tables. – 2, chair. – , annexes. – 1, references. – 128.

Keywords: mining, data, tool, integration, confidentiality, processing, healthcare.

The qualification work is devoted to the study of methods and means of analytical processing of information on COVID-19. In the first, the state of the subject area is recorded. Data extraction for COVID-19 was analyzed. The stages of their preliminary processing are highlighted. The second section examines the methods of extracting information on COVID-19. The collection and preliminary processing of social information on COVID-19 is considered. Peculiarities of data integration on COVID-19 are analyzed. Describes tools for extracting information about COVID-19. The third section of the qualification work examines the merging and extraction of information on COVID-19 obtained from IoH sources. The use of artificial intelligence to process information on COVID-19 is described. Methods of teaching artificial intelligence systems in diagnosing COVID-19 are analyzed. The section "Occupational health and safety in emergencies" describes the regulations that define the actions of the employer, whose employee fell ill with COVID-19 and increase the resilience of economic activities in wartime.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

CNN (англ. Convolutional Neural Network) – Нейромережа зворотнього поширення.

EHRs (англ. Electronic Health Record Systems) – Електронні системи обліку здоров'я.

EMR (англ. Electronic Medical Record) – Електронна медична карта.

IoH (англ. Internet of Health Things) – Інтернет речей для здоров'я.

GUI (англ. Graphical user interface) – Графічний інтерфейс користувача.

KNIME (нім. Konstanz Information Miner) – Констанс інформаційний майнер.

LSH (англ. Locality-sensitive hashing) – Локально-чутливе хешування.

LDA (англ. Latent Dirichlet allocation) – Прихований розподіл Діріхле.

MLNN (англ. Multi-Layer Neural Network) – Багатошарова нейронна мережа.

NN (англ. Neural Network) – Нейронна мережа.

PDFM (англ. Privacy-free Data Fusion and Mining) – Злиття та видобування даних без конфіденційності.

PPV (англ. Positive Predicted Value) – Позитивна прогнозована цінність.

SVM (англ. Support Vector Machine) – Метод опорних векторів.

ГМ – Графічна модель.

ЕМК – Електронна медична картка.

КТ – Комп'ютерна томограма.

ШІ – Штучний інтелект.

ЗМІСТ

ВСТУП	11
1 АКТУАЛЬНІСТЬ АНАЛІТИЧНОГО ОПРАЦЮВАННЯ	
ВІДОМОСТЕЙ ЩОДО COVID-19	13
1.1 Стан предметної області	13
1.2 Видобування даних щодо COVID-19	19
1.3 Попереднє опрацювання відомостей щодо COVID-19	21
1.4 Висновок до першого розділу	24
2 ТЕХНІКИ ВИДОБУВАННЯ ДАНИХ ЩОДО COVID-19	25
2.1 Методи видобування відомостей щодо COVID-19.....	25
2.1.1 Класифікація даних щодо COVID-19	26
2.1.2 Дерева рішень при опрацюванні відомостей щодо COVID-19 ...	26
2.1.3 Використання нейронних мереж для опрацювання відомостей щодо COVID-19.....	27
2.1.4 Логістична регресія та метод К-найближчих сусідів (K-NN) при опрацюванні даних щодо COVID-19	28
2.1.5 Підтримка векторної машини (SVM) при опрацюванні відомостей щодо COVID-19	29
2.1.6 Генетичні алгоритми при опрацюванні даних щодо COVID-19.	29
2.1.7 Мережа Байєса при опрацюванні відомостей щодо COVID-19 ...	30
2.1.8 Апріорі-алгоритм	31
2.1.9 Кластеризація даних щодо COVID-19	31
2.1.10 Асоціація відомостей щодо COVID-19.....	32
2.1.11 Прихований розподіл Діріхле	33
2.2 Збирання та попереднє опрацювання соціальних відомостей щодо COVID-19	34
2.3 Особливості інтеграції даних щодо COVID-19.....	35
2.3.1 Шифрування COVID-19 даних	35

	10
2.3.2 Забезпечення конфіденційності відомостей щодо COVID-19	36
2.3.3 Анонімізація даних	37
2.4 Інструменти для видобування відомостей щодо COVID-19	38
2.5 Висновок до другого розділу	40
3 ПРАКТИКИ ВИКОРИСТАННЯ АНАЛІТИЧНИХ ІНСТРУМЕНТІВ ДЛЯ ОПРАЦЮВАННЯ ВІДОМОСТЕЙ ЩОДО COVID-19	41
3.1 Злиття та видобування відомостей щодо COVID-19 отриманих з ІоН-джерел.....	41
3.2 Використання штучного інтелекту для опрацювання відомостей щодо COVID-19	48
3.3 Методи навчання систем ШІ при діагностуванні COVID-19	51
3.3.1 Контрольоване навчання	51
3.3.2 Неконтрольоване навчання	53
3.3.3 Підкріплення навчання	54
3.4 Порівняльна характеристика методів навчання систем ШІ.....	56
3.5 Методи глибинного навчання та COVID-19	57
3.6 Висновок до третього розділу	59
4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ	60
4.1 Аналіз нормативних документів, які визначають дії роботодавця, працівник якого захворів на COVID-19	60
4.2 Підвищення стійкості роботи об'єктів господарської діяльності у воєнний час.....	63
ВИСНОВКИ.....	65
ПЕРЕЛІК ДЖЕРЕЛ	66
ДОДАТКИ	

ВСТУП

Актуальність теми. SARS-CoV-2 (коронавірус 2019, COVID-19) створив виняткову ситуацію у 2020 році, хоча багато інших вірусів, що з'явилися в попередні роки не спричинили настільки глобальної надзвичайної ситуації. Пандемія COVID-19 кардинально змінила поведінку жителів та умови їх життя у всіх куточках землі. У період карантинних заходів пов'язаних з COVID-19 спостерігається значне збільшення обсягів медичних даних порівняно з аналогічними періодами попередніх років. Пандемія COVID-19 змінює моделі аналітичного опрацювання медичних даних та підходи до їх формування. Тому методи та засоби аналітичного опрацювання відомостей щодо COVID-19 є актуальним напрямком сучасних наукових досліджень.

Мета і задачі дослідження. Метою кваліфікаційної роботи є підвищення ступеня повноти подання інформації щодо COVID-19. Для досягнення поставленої мети слід було виконати наступні завдання:

- Зафіксувати стан досліджень в даній предметній області.
- Проаналізувати відомі методи видобування відомостей щодо COVID-19.
- Дослідити особливості інтеграції даних щодо COVID-19.
- Розробити практичні рекомендації для видобування відомостей щодо COVID-19.

Об'єкт дослідження: процеси аналітичного опрацювання відомостей щодо COVID-19.

Предмет дослідження: методи аналітичного опрацювання відомостей щодо COVID-19 та засоби їх практичної реалізації.

Наукова новизна одержаних результатів кваліфікаційної роботи полягає у тому, що отримав подальший розвиток метод злиття та видобування даних отриманих з використанням ІоН-джерел. Сформовано

порівняльну характеристику методів навчання систем ШІ для їх використання при опрацюванні даних щодо COVID-19.

Практичне значення одержаних результатів. Реалізований в кваліфікаційній роботі метод, адаптовано для знеособлення персональних даних пацієнтів, інфікованих COVID-19.

Апробація результатів кваліфікаційної роботи. Основні результати проведених досліджень обговорювались на VIII науково-технічній конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 2020 р.).

Публікації. Основні результати кваліфікаційної роботи опубліковано у двох працях конференції (Див. додаток А).

Структура й обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з 128 найменувань та 1 додатку. Загальний обсяг кваліфікаційної роботи складає 85 сторінок, з них 54 сторінки основного тексту, який містить 8 рисунків та 2 таблиці.

1 АКТУАЛЬНІСТЬ АНАЛІТИЧНОГО ОПРАЦЮВАННЯ ВІДОМОСТЕЙ ЩОДО COVID-19

1.1 Стан предметної області

Галузь охорони здоров'я зазнає великих змін внаслідок пандемії COVID-19 та, як наслідок, еволюції в накопиченні величезних обсягів медичних даних. Така ситуація спричиняє розвиток мультидисциплінарних досліджень. Велика кількість вітчизняних та світових вчених аналізують та досліджують дані галузі охорони здоров'я зібрані щодо COVID-19, яка є активною міждисциплінарною областю видобування даних. Технологічний прогрес спричиняє експоненційне зростання відомостей щодо охорони здоров'я. Оцифрування медичних записів щодо COVID-19 призвело до швидкого зростання інформаційно-технологічного сектору охорони здоров'я. Електронні системи медичних записів (EHR) – це сховища даних, які є оцифрованим форматом для зберігання медичних даних [1]. Сектор охорони здоров'я управляє величезними обсягами даних щодо COVID-19, які потрібно проаналізувати з метою формування оптимальних рекомендацій для забезпечення покращення процесу прийняття рішень. Основна проблема полягає у використанні методів видобування даних для ефективного виявлення корисної та важливої інформації серед великих обсягів доступних щодо COVID-19 відомостей. Розроблення та формування нових методів є важливим для підвищення ефективності опрацювання великих за обсягом даних щодо COVID-19 у галузі охорони здоров'я. У даній кваліфікаційній роботі освітнього рівня «магістр» зібрано відомості, які демонструють важливість видобування даних щодо COVID-19 зібраних у галузі охорони здоров'я та соціальних джерел. Проведені дослідження головним чином зосереджують увагу на необхідності аналізу даних щодо COVID-19 та їх застосуванні в секторі охорони здоров'я, різних методах прогнозування та

описового аналізу даних, які можуть бути використані при опрацюванні відомостей щодо COVID-19.

З появою концепції Інтернету здоров'я (IoH) традиційні медичні та здорові послуги поступово переходять до Інтернету, генеруючи при цьому значну кількість медичних даних, пов'язаних із пацієнтами, лікарями, медициною, медичною інфраструктурою, тощо, включаючи відомості щодо COVID-19. Ефективне злиття та аналіз IoH-даних мають позитивне значення для наукової діагностики катастроф на кшталт COVID-19 та підвищення якості надання медичної допомоги. Однак IoH-дані часто розподіляються між різними відділами та містять частку конфіденційних відомостей громадян та користувачів. Це в свою чергу підвищує складність завдання ефективного інтегрування та видобування відомостей щодо COVID-19 в т.ч. зібраних з використанням IoH. При цьому слід забезпечити конфіденційні відомості громадян та користувачів від розголошення. Щоб подолати зазначені складності, слід використовувати інформаційно-технологічні рішення для інтеграції та видобування медичних даних щодо COVID-19 отримані із багатьох різнотипових інформаційних джерел для покращення послуг охорони здоров'я застосовують PDFM (Злиття та видобування даних без конфіденційності) [2]. За допомогою PDFM можна шукати подібні медичні записи щодо COVID-19 у потрібний момент часу та в умовах збереження конфіденційності, щоб запропонувати пацієнтам кращі та ефективніші медичні послуги.

В умовах пандемії COVID-19 та завдяки постійному розвитку інформаційних та комунікаційних технологій, що супроводжується поступовим впровадженням цифрового програмного забезпечення в сфері охорони здоров'я, різні медичні підрозділи та установи накопичують значну кількість історичних даних (наприклад, медичні карти пацієнтів, ефективні рішення щодо лікування тощо). Ці відомості є основним джерелом великих за обсягом даних для «Інтернету охорони здоров'я» (IoH) [3]. Ступінь

використання таких даних ІоН в умовах пандемії COVID-19 є ключовим критерієм для оцінювання поточної епідеміологічної ситуації та кількісного оцінювання рівня поінформованості медичних установ, підрозділів чи відділень [4].

Як правило, більшість історичних записів даних ІоН щодо COVID-19 містять важливу для медичних підрозділів та установ інформацію. Це, зокрема, відомості про історію хвороб пацієнтів в певні періоди часу. Видобування та аналіз таких історичних записів ІоН-даних може значно сприяти науковому та обґрунтованому діагностуванню симптомів COVID-19 та прийняттю рішень лікарями, а також прогнозуванню тенденцій розвитку епідеміологічної ситуації та уникненню катастрофічних наслідків [5]. Отже, є надзвичайна необхідність збирати, інтегрувати, об'єднувати та аналізувати мультиджерельні записи даних щодо COVID-19 для надання високоякісних медичних послуг, доступних для обширного кола громадян.

Однак історичні записи даних щодо COVID-19, отримані від пацієнтів часто містять конфіденційні персональні відомості про пацієнта (наприклад, артеріальний тиск, температура), оскільки багато пацієнтів часто не хочуть повідомляти іншим громадянам відомості про свій стан здоров'я [6]. Таким чином, пацієнти або зацікавлені сторони історичних записів даних щодо COVID-19 не наважуються розголошувати свої записи ІоН-даних широкому загалу. Крім того, відсутні стимули для обміну персоніфікованими відомостями щодо COVID-19 з іншими установами, в тому числі не медичними. Вищезазначені два питання суттєво обмежують використання історичних записів щодо COVID-19. Як наслідок, не зважаючи на те що багато лікарень, інших медичних та не медичних установ накопичили значну кількість історичних записів даних щодо COVID-19, вони рідко передають дані назовні через проблеми конфіденційності. Крім того, історичні записи даних щодо COVID-19 часто поширюються між різними інформаційно-

технологічними платформами чи установами, інтеграція та злиття яких ще більше посилює проблеми розкриття конфіденційності.

Враховуючи вищезазначені проблеми в галузі медичних даних щодо COVID-19, доцільне використовуємо хеш-методів, для реалізації захисту приватних даних під час використання та аналізу. Оскільки хеш-методи є односпрямованими способами картографування даних, ціль захисту конфіденційності може бути досягнута.

Основним обмеженням видобування даних щодо COVID-19 у галузі охорони здоров'я є неоднорідні та об'ємні вихідні дані. Це співвідноситься з даними щодо COVID-19 з різних джерел, таких як прийом лікаря, лабораторні обстеження, аналізи та обстеження лікаря тощо. Завдяки цьому доступність даних щодо COVID-19 може бути обмежена, а процес ускладнюється для збору, зберігання та аналізу даних. Однак не слід ігнорувати будь-які дані щодо COVID-19, оскільки всі компоненти даних можуть мати значний вплив на діагноз та прогресування хвороби пацієнтів. Тому дані потрібно збирати. Іншим питанням є неповні або нестандартні дані щодо COVID-19, неточні або відсутні дані в медичній документації. Наприклад, можна використовувати різні формати для збору фрагментів даних щодо COVID-19 у різних джерелах. Без звичайної клінічної термінології аналіз даних щодо COVID-19 у секторі охорони здоров'я також надзвичайно складний. Подальшою перешкодою для ефективного видобування даних щодо COVID-19 є погані математичні характеристики та неканонічний характер таких великих обсягів, складних та неоднорідних даних. Існують також інші важливі питання медичних даних щодо COVID-19, такі як контроль даних, етичні проблеми, соціальні та правові проблеми тощо. Інша проблема полягає в тому, що результати аналізу даних щодо COVID-19 можуть виявити різні важливі та цікаві тенденції, які можуть бути неактуальними через величину обсягів даних. Ще однією вимогою до успішного застосування інтелектуального аналізу даних щодо COVID-19 є

знання в області домену, а також належне розуміння методів інтелектуального аналізу даних. Крім того, потрібні значні витрати часу, ресурсів та зусиль для вдосконалення інформаційних технологій аналізу даних. Введення даних щодо COVID-19 має бути систематичним та належним чином зберігатись для подальшого використання. Основною вимогою є ретельне планування, напрацювання щодо інформаційно-технологічної підготовки, обізнаність про ефективність інформаційних технологій, їх використання та спільна робота усіма, хто займається видобуванням даних щодо COVID-19.

Важливим аспектом дослідження є психічний стан інфікованих COVID-19 пацієнтів та населення загалом. Спалах COVID-19 змусив людей змінити своє звичне життя та практикувати соціальну дистанцію. Така раптова зміна може різко підвищити рівень стресу людей та призвести до інших проблем психічного здоров'я. Труднощі, спричинені спалахом COVID-19 у різних географічних регіонах, можуть визначити причину та ступінь стресу у людей, що відповідає ризику розвитку серйозної депресії [7]. Згідно з опитуванням [8], майже половина дорослих людей у Сполучених Штатах повідомили, що їх психічне здоров'я негативно змінилося через хвилювання та стрес через COVID-19. З продовженням пандемії COVID-19, зростатиме навантаження на психічне здоров'я, оскільки почуття нормальності людей продовжує порушуватися через соціальне дистанціювання, закриття бізнесу, шкіл та громадських місць. Постійна невизначеність та брак ресурсів призводять до ще більшої ізоляції та фінансових труднощів.

Географічні інформаційні системи (ГІС) та видобування даних у соціальних мережах стали важливими інструментами вивчення просторового розподілу інфекційних захворювань [9], [10], [11] і можуть бути використані для дослідження просторово-часової картини психічного стресу, спричиненого пандемією COVID-19. Зокрема, дані соціальних мереж

(наприклад, дані Twitter) дають унікальну можливість дізнатись про настрої, почуття та поведінку громадян, які відображають їхній психо-емоційний стан [12], [13]. Багато наукових публікацій зосереджувались на використанні підходів, що базуються на особливостях аналізу настроїв та емоційного аналізу з використанням даних Twitter [14], [15], [16]. Зокрема, Го та його колеги [17] досліджували використання уніграм, біграм та їх поєднання у навчанні класифікаторів для аналізу настроїв твітів. Були навчені різні контрольовані класифікатори, включаючи максимальну ентропію, наївний Байєс [18], класифікатори з підтримкою векторних машин (SVM) та їхню ефективність на n-грамах. Однак деякі методи, що використовувались раніше, застаріли; наприклад, вони не враховували смайлики для формування індексу настрою, але сьогодні багато користувачів Twitter частіше використовують смайлики [19]. Барбоза та Фенг [20] показали, що n-грами не є корисними для класифікації твітів, оскільки невикористані слова в твітах можуть спричинити проблеми під час навчання класифікатора. Пак та Парубек [21] запропонували використовувати такі функції мікроблогів, як хештеги, смайлики, повторні твіти та коментарі для підготовки SVM-класифікатора, і показали, що це призвело до вищої точності, ніж тренування з використанням n-грам. У кількох статтях розглядається ефект від використання функцій тегу Part-Of-Speech (POS) у класифікаторах тексту [22]. Абаді та його колеги досліджували функції POS, лексики та мікроблогів. Результати показали, що найбільш релевантними є ті функції, які поєднують попередню полярність із POS-мітками слів. Однак є неоднозначні результати щодо використання POS-тегів. Го та його колеги показали, що POS-теги спричиняють зниження продуктивності, хоча POS-теги можуть бути ефективними показниками емоцій у тексті та слугувати корисною функцією в аналізі думок чи настроїв щодо COVID-19. Більше того, підходи до початкового завантаження, які покладаються на перелік слів думок чи емоцій, щоб знайти інші такі слова у великому тексті, стають все

більш популярними та довели свою ефективність [23], [24]. Міхальчеа, Банеа та Вейб [25] описали два типи методів для розподілу лексики суб'єктивності на словникові та корпусні. Їх дослідження розпочалися з невеликого набору набраних суб'єктивних слів, що склалися вручну, та за допомогою Інтернет-словника створили більший лексикон слів потенційних кандидатів. Подібна модель завантаження була ефективно використана для побудови системи аналізу настроїв для формування огляду стану здоров'я щодо наркотиків та ліків, створеного користувачами [26]. Однак усі вищезазначені методи лише виявляють загальну емоцію твітів і не мають можливості детально моделювати рівні депресії, які будуть наслідком COVID-19.

1.2 Видобування даних щодо COVID-19

Видобування даних – це процес оцінки баз та сховищ даних, щоб отримати з них нові уявлення та знання. Видобування даних стає дедалі популярнішим у галузі охорони здоров'я, особливо це стосується даних щодо COVID-19. Це містить великий потенціал як для галузі охорони здоров'я загалом так і в контексті COVID-19, для того, щоб дати можливість інституціям охорони здоров'я системно використовувати дані та аналітичні засоби для виявлення ефективних та найкращих практик, що покращують якість медичних послуг та зменшують фінансові витрати [27]. Завдяки експоненціальному збільшенню кількості електронних медичних записів щодо COVID-19, видобування даних має великий потенціал для галузі охорони здоров'я. Раніше лікарі, медичні заклади та установи зберігали інформацію про пацієнтів у фізичних документах, що було досить складно. Цифровізація та впровадження нових інформаційних технологій в медичній галузі зменшує людські зусилля та полегшує процедури аналізу даних.

Видобування даних щодо COVID-19 перекроює багато галузей, включаючи сферу охорони здоров'я. Програми для видобування даних

можуть принести користь усім працівникам, задіяним у сфері охорони здоров'я. Структура даних щодо COVID-19 спрощує та автоматизує робочі процеси організацій охорони здоров'я. Інтеграція інтелектуального аналізу даних у структуру даних щодо COVID-19 зменшує зусилля лікарів у прийнятті рішень та забезпечує нові цінні медичні знання. Електронні медичні картки (ЕМК) є поширеними у закладах охорони здоров'я. Завдяки поліпшеному доступу до великого обсягу даних щодо COVID-19, що включають відомості про пацієнтів, медичні працівники зараз зосереджуються на підвищенні ефективності та послідовності використання інтелектуального аналізу даних у своїх організаціях. Прогнозні моделі надають медичним працівникам найкращу інформаційну підтримку та знання щодо COVID-19 [28]. Метою інтелектуального аналізу даних щодо COVID-19 є формування ефективних моделей прогнозування, забезпечення надійних прогнозів, підтримка лікарів у вдосконаленні діагностики та плануванні лікування тощо [29]. Інтелектуальний аналіз даних щодо COVID-19 може допомогти лікарям визначити найкращі напрямки дій, мінімізувати випадки невідомих реакцій на ліки, підвищити якість та безпеку пацієнтів, визначити фактори, пов'язані з шахрайством у медичному страхуванні, підібрати спеціалістів до потреб пацієнтів тощо. Інтелектуальний аналіз даних щодо COVID-19 допомагає організаціям охорони здоров'я оцінити ефективність лікування, рятує життя пацієнтів за допомогою інтелектуальної медицини, керує стосунками із клієнтами, виявляти шахрайство та зловживання та у багатьох інших сферах.

Обсяг даних щодо COVID-19, що виробляються у сфері охорони здоров'я, повинен бути перетворений на корисні знання для прийняття рішень. Видобування даних щодо COVID-19 – це велика перспектива в галузі охорони здоров'я, яка аналізує складність даних для отримання інформації [30]. Процес аналізу даних щодо COVID-19 допомагає виявити корисні

відомості від етапу відбору до етапу виявлення знань. На рисунку 1.1 подано ключові етапи процесу видобування даних у медичній галузі.

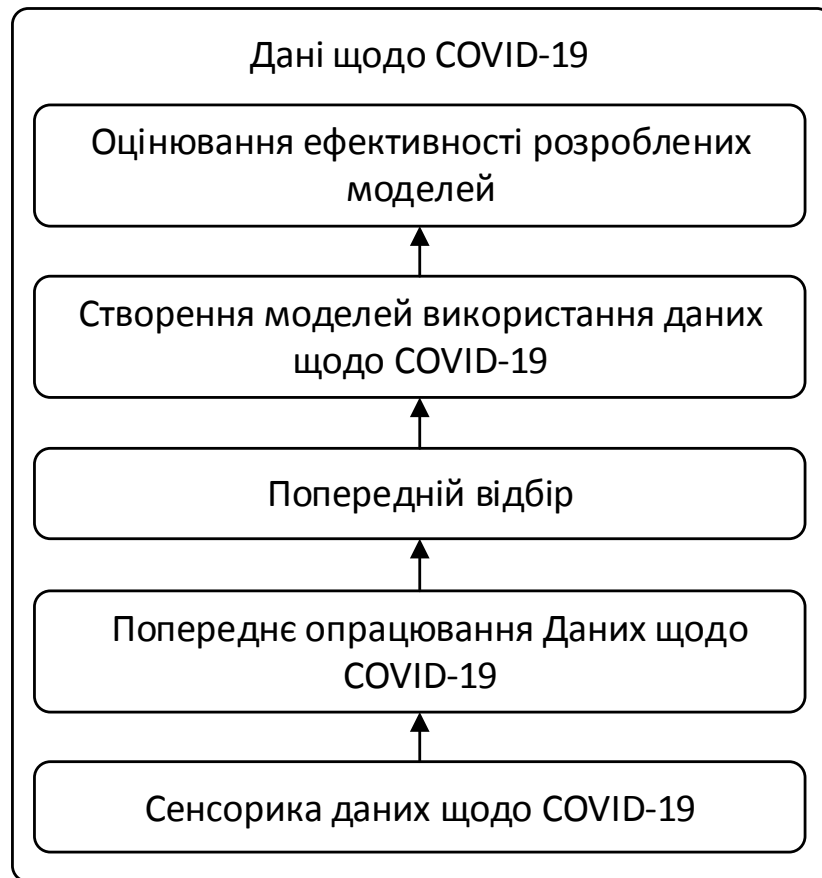


Рисунок 1.1 – Процес опрацювання даних щодо COVID-19

Нижче описано ключові етапи процесу аналітичного опрацювання даних щодо COVID-19 для формування різноманітних моделей та оцінювання їх ефективності.

1.3 Попереднє опрацювання відомостей щодо COVID-19

Техніка попереднього опрацювання відомостей щодо COVID-19 використовується при видобуванні даних для перетворення вихідних даних у корисні та ефективні знання. Етапи попереднього опрацювання даних включають:

- очищення відомостей щодо COVID-19;
- створення даних щодо COVID-19;
- зменшення даних щодо COVID-19.

Дані щодо COVID-19 можуть мати кілька розділів, які є незначними або відсутніми. Очищення даних щодо COVID-19 з використанням аналітичного програмного забезпечення виконується для їх опрацювання від відомостей, що викликають шум, відсутність даних, тощо. Перетворення даних щодо COVID-19 здійснюється для їх трансформації у відповідну форму, придатну для процесу видобування [31]. Сюди входить нормалізація даних щодо COVID-19, збирання атрибутів, дискретизація та генерація ієрархії даних. Видобування даних щодо COVID-19 – це техніка, що використовується для управління великими обсягами даних. У цьому випадку аналіз даних щодо COVID-19 стає важчим при роботі з великими обсягами даних. Щоб позбутися цього недоліку, використовується техніка зменшення даних щодо COVID-19. Зазначена техніка спрямована на підвищення ефективності зберігання даних, зниження вартості їх зберігання та аналізу відомостей щодо COVID-19. Вона складається з агрегування кубів даних, збору атрибутів підмножини, зменшення нумерації та зменшення розмірності.

Вибір функціональних особливостей можна визначити як вибір мінімального підмножини функцій, які насправді необхідні для будь-якого процесу видобування даних щодо COVID-19. При цьому набір функцій може бути завеликим, а його ефективність може бути занижена. Крім того, вибір функції мінімізує кількість основних функціональних наборів, необхідних для оптимізації точності моделей, що в свою чергу допомагає зменшити простір, необхідний функціональним наборам. Це також усуває надмірні шуми, які можуть бути притаманні функціональним наборам, що також покращує ефективність алгоритмів інтелектуального аналізу даних щодо COVID-19. Метою вибору ознак є розроблення ефективних моделей.

Вибір ознак в процесі опрацювання відомостей щодо COVID-19 складається з чотирьох етапів:

- розроблення підмножини;
- оцінювання підмножини;
- визначення критеріїв відбору;
- остаточний опис ознак підмножини.

Функціональний набір перевіряється на першому етапі після усунення невідповідностей, таких як нульові значення та надмірності відомостей щодо COVID-19. Після пошуку функціонального набору починається процес генерації підмножини. Оцінювач атрибутів здійснює оцінку сформованої підмножини. Процес генерації та оцінки підмножини триває до тих пір, поки не будуть дотримані критерії відбору. Остаточний функціональний набір підмножини відомостей щодо COVID-19 формується лише після завершення вищевказаного процесу.

Модель видобування відомостей щодо COVID-19 отримує дані з структури видобутку, а потім аналізує дані за допомогою алгоритмів видобування даних. Структура майнінгу відомостей щодо COVID-19 зберігає інформацію, яка визначає джерело даних. Модель видобування відомостей щодо COVID-19 зберігає дані статистичного процесу, такі як закономірності, знайдені в результаті аналізу.

Кожен тип моделей даних щодо COVID-19 створює різний набір шаблонів, наборів елементів, правил або формул, які можна використовувати для прогнозування. Алгоритмами, які можуть бути використані в процесі розроблення моделей даних щодо COVID-19, є: дерево рішень, нейронні мережі та логістична регресія, тощо. Існують різні підходи до оцінювання роботи моделей. Зазвичай використовуються відповідні критерії вимірювання, зокрема: точність, чутливість, специфічність та F-міра. Точність визначається як співвідношення правильно класифікованих випадків. Чутливість вимірює співвідношення фактичних позитивів, які були

правильно визначені. Специфічність вимірює співвідношення правильно визначених фактичних негативів. Точність, також відома як позитивна прогнозована цінність (PPV), вимірює співвідношення справжніх позитивних результатів до передбачуваних позитивних випадків COVID-19. F-міра – це гармонічне середнє значення точності та відклику.

1.4 Висновок до першого розділу

В першому розділі кваліфікаційної роботи подано аналіз стану предметної області видобування даних щодо COVID-19. Наведено узагальнений опис процесів видобування даних щодо COVID-19. Описано особливості та етапи переднього опрацювання відомостей щодо COVID-19.

2 ТЕХНІКИ ВИДОБУВАННЯ ДАНИХ ЩОДО COVID-19

2.1 Методи видобування відомостей щодо COVID-19

Методи видобування даних щодо COVID-19 бувають двох типів, описові та передбачувальні. Описовий аналіз використовується для видобування даних щодо COVID-19 та надання актуальної інформації про минулі та останні події [32]. Прогнозний аналіз даних щодо COVID-19 дає відповіді на перспективні запити, які переходять із використанням історичних даних як головного принципу прийняття рішень [33].

На рисунку 2.1 подано класифікацію методів аналізу даних в медичній галузі, які можуть бути використані для аналізу відомостей щодо COVID-19.

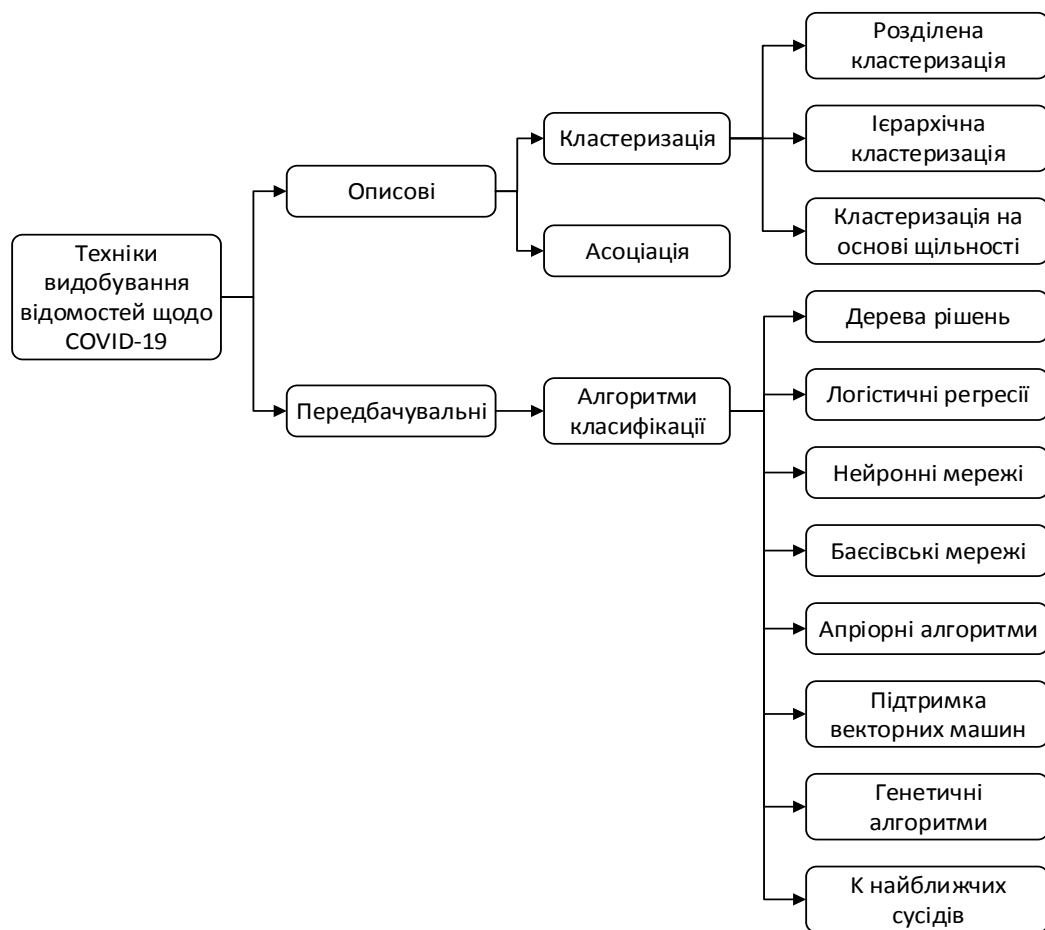


Рисунок 2.1 – Методи видобування даних застосовні для опрацювання відомостей щодо COVID-19

2.1.1 Класифікація даних щодо COVID-19

Класифікація – це загальноприйнята техніка видобування даних щодо COVID-19 і використовується для класифікації кожного елемента в один із заздалегідь визначеного наборів класів або груп у наборі даних. Метод класифікації використовує різноманітні математичні методи, зокрема, дерева рішень, нейронні мережі, логістичні регресії, векторні машини, генетичні алгоритми, мережі Байєса.

Програмне забезпечення для класифікації даних щодо COVID-19 може бути навчене на основі набору даних для прогнозування майбутніх подій [34]. Особливості наборів даних можна класифікувати як низькі, помірні, високі та дуже високі щодо класифікації на основі симптомів COVID-19 та інших діагностованих захворювань. Класифікація є найбільш часто використовуваним методом ідентифікації, прогнозування та оптимізації у галузі охорони здоров'я та зокрема COVID-19.

2.1.2 Дерева рішень при опрацюванні відомостей щодо COVID-19

Дерева рішень – це широко застосовувана техніка опрацювання даних щодо COVID-19, оскільки користувачі можуть легко зрозуміти виявлені закономірності. У зазначеній техніці просте запитання чи умова, що має складені відповіді, є коренем дерева рішень [35]. Тоді кожна відповідь приводить до групи запитань або умов, які допомагають визначити потрібні відомості щодо COVID-19, щоб остаточне рішення могло ґрунтуватися на них. Це деревоподібна модель даних у базах та сховищах даних. Прийняття рішень використовується для прогнозування можливих подій в плані розгортання епідеміологічної ситуації COVID-19 та сприяє підвищенню точності результатів. Існують кінцеві та нетермінальні вузли дерева рішень. Кожен нетермінальний вузол елемента даних являє собою тест або умову.

Дерева рішень використовуються для класифікації екземплярів, їх сортування відповідно до термінальних вузлів з нетермінальних. Вибір потрібної гілки повністю залежить від результатів тестування. Дерева рішень зазвичай використовуються при аналізі досліджень операцій для обчислення умовних ймовірностей. За підтримки дерев рішень можна вибрати найкращі альтернативи, і на основі максимального набору знань щодо обходу від кореневого до листового вузла сформуванати унікальне розділення класів відомостей щодо COVID-19.

2.1.3 Використання нейронних мереж для опрацювання відомостей щодо COVID-19

Нейронні мережі є інтелектуальними сутностями для видобування відомостей щодо COVID-19, оскільки вони функціонально організовані подібно до людського мозку і націлені на пошук прихованих зв'язків між даними щодо COVID-19. Штучний нейрон – це блок опрацювання даних, який отримує зважені вхідні значення від наборів атрибутів, перетворює отримане значення за базовою формулою і передає результат у формі виходу до інших атрибутів. Нейронна мережа є найкращим алгоритмом для класифікації дерев рішень та підтримки векторних машин. Основна мета використання нейронних мереж – розпізнавати закономірності та виконувати класифікаційні завдання при опрацюванні відомостей щодо COVID-19. При регулюванні ваг, це допомагає мінімізувати помилки через адаптивний характер. Штучні нейрони працюють разом, щоб комплексно генерувати вихідну функцію. На етапі навчання нейронна мережа повинна навчитися передбачати правильні мітки класів вхідних даних щодо COVID-19, змінюючи ваги. Нейронні мережі мають перевагу, оскільки на відміну від простих методів моделювання вони можуть передбачати нелінійні взаємозв'язки відомостей щодо COVID-19. Нейронні мережі допомагають

аналізувати медичні дані щодо COVID-19 та використовуються як один з найпопулярніших алгоритмів опрацювання даних у медичній галузі, зокрема відомостей щодо COVID-19. Штучні нейронні мережі у медичній галузі використовується для класифікації тканин, діагностуванні захворювань, зокрема COVID-19. Багатошарова нейронна мережа (MLNN) використовує приховані шари, за допомогою яких вирішує проблему класифікації нелінійних множин. Зазвичай ці приховані шари інтерпретуються як гіперплощини. Такі нейронні мережі використовуються для класифікації різних категорій даних щодо COVID-19.

2.1.4 Логістична регресія та метод К-найближчих сусідів (K-NN) при опрацюванні даних щодо COVID-19

При використанні логістичної регресії для опрацюванні відомостей щодо COVID-19, атрибут «label» передбачається на основі значень вхідних атрибутів. Він описує відношення між атрибутом «label» та набором вхідних атрибутів, що його описують. У галузі охорони здоров'я логістична регресія використовується для прогнозування та ідентифікації захворювань, зокрема COVID-19. Це, насамперед, статистичний інструмент, який використовується для видобування даних щодо COVID-19. Він аналізує лише логістичні та нелінійні категоріальні дані щодо COVID-19.

Метод К-найближчих сусідів (K-NN) використовує методи класифікації та регресії і є простим та ефективним у використанні інструментом. У K-NN нові дані щодо COVID-19, що вводяться в базу даних, аналізуються шляхом знаходження підмножини набору даних для отримання оптимального рішення та прогнозування точного результату. Ця методика використовується як критерій прогнозування та виявлення COVID-19.

2.1.5 Підтримка векторної машини (SVM) при опрацюванні відомостей щодо COVID-19

Підтримка векторної машини (SVM) застосовується за допомогою класифікації та регресії під час контрольованого навчання для оцінки даних щодо COVID-19. Він ділиться на два класи гіперплощини. SVM автоматизує процеси опрацювання відомостей щодо COVID-19, роблячи їх більш ефективними. Метод обгрунтовано застосовується в галузі охорони здоров'я для виявлення прогностичних особливостей COVID-19. Гіперплощина – це розподіл між двома результатами в двійковій класифікаційній задачі, такі як прогнозування ускладнень від COVID-19. Ключовим завданням використання гіперплощин є максимізація поділу між точками даних. Для зашумлених даних похибка зводиться до мінімуму, максимізуючи поле між двома окремими класами прикладів та визначаючи гіперплощину як центральну лінію розділового простору. Існують дві форми SVM. Перший – це лінійні SVM, які розділяють точки даних, використовуючи лінійну межу для прийняття рішення. Метод ефективно використовується для наборів відомостей щодо COVID-19, які можна легко розділити на дві частини. Складні набори даних важко визначити за допомогою лінійного ядра, яке використовує другу форму SVM. Для цього використовуються нелінійні SVM, що розділяють набори даних з використанням нелінійних меж рішення. SVM показує точність у проблемах двійкової класифікації відомостей щодо COVID-19.

2.1.6 Генетичні алгоритми при опрацюванні даних щодо COVID-19

Генетичний алгоритм – це генетична та відбірна методика пошуку та оптимізації відомостей щодо COVID-19. Генетичні алгоритми використовуються насамперед у нейронних наборах, які виконують роль

керівництва для процесу навчання алгоритмів інтелектуального аналізу відомостей щодо COVID-19, а не пошуку шаблонів. Вони часто використовуються для формулювання гіпотези про змінні та залежності між ними у вигляді правил асоціації іншого формалізму в інтелектуальному аналізі даних COVID-19. У генетичному алгоритмі існує популяція, складена з багатьох особин, яка еволюціонує до стану, коли придатність максимізується відповідно до певних правил відбору. Початкова сукупність правил створюється випадковим чином, причому кожне правило являє собою вирішення проблеми. Натомість пари правил обираються батьками, які, як правило, є найсильнішими правилами. Генетичний алгоритм складається по суті з трьох операторів – відбору, кросоверу та мутації. При виборі підбирається підходяща струна на основі придатності для розведення нового покоління, потім поєднуються придатні струни для отримання кращого потомства. Мутація змінює ланцюжок локально, щоб зберегти генетичне. Для припинення роботи алгоритму сукупність оцінюється в кожному поколінні. Після чого відбувається повторне оцінювання відомостей щодо COVID-19.

2.1.7 Мережа Байєса при опрацюванні відомостей щодо COVID-19

Мережа Байєса – це специфічна форма мережі, яка представляє невизначену інформацію про домен. Належить до категорії графічних імовірнісних моделей (ГМ). Вузли в мережі Байєса представляють змінні, а конкретні ребра – імовірнісні залежності. Для кожної змінної мережа Байєса визначає два типи знань щодо COVID-19. У медичній науці класифікатор Байєса базується на теорії ймовірності і може бути використаний як логічний процес проведення медичної діагностики, особливо в автоматизованих системах підтримки прийняття рішень щодо COVID-19.

2.1.8 Апріорі-алгоритм

Це алгоритм для вивчення часкового вибору наборів елементів та правил асоціації над реляційними базами даних з відомостями щодо COVID-19. Це продовжується шляхом визначення звичайних окремих сутностей у базі даних та їх розширення до збільшуваних наборів елементів, якщо ці набори елементів з'являються в базі даних щодо COVID-19 досить часто. Це вирішує задачі від найскладніших до найменших та групує відомості щодо COVID-19, які потрібно проаналізувати. Як тільки всі задачі будуть вирішені, алгоритм зупиняється. Він показує взаємозв'язок між двома входами для розрізнення послідовних і непослідовних входів. Для прогнозування захворювань в галузі охорони здоров'я, зокрема COVID-19, можна використовувати різні типи апріорі-алгоритмів, наприклад Хеш-таблиці, скорочення транзакцій, розподіл розділів тощо.

2.1.9 Кластеризація даних щодо COVID-19

Процес кластеризації використовує автоматизовану техніку для побудови кластера наділених подібними характеристиками сутностей [36]. Це дозволяє виконувати процедури визначення класів та розміщення об'єктів у кожному класі. У кінцевому наборі відомостей щодо COVID-19 відбувається ідентифікація категорії даних яка використовується для прогнозування потрібного результату. Kmeans та xmeans – це алгоритми, що використовуються для оцінки результатів аналітичного опрацювання відомостей щодо COVID-19. Кластеризацію можна розділити:

– Кластеризація за допомогою сегментів – дозволяє класифікувати пов'язані ознаки набору відомостей щодо COVID-19 за різними групами та аналітиками, використовується для пошуку згенерованих кластерів.

Розподілений алгоритм кластеризації допомагає прогнозувати та діагностувати симптоми для конкретних захворювань, зокрема COVID-19.

– Ієрархічна кластеризація використовується для кластеризування функції набору даних в ієрархічній формі, яка використовується для формування прогностичних моделей в галузі охорони здоров'я, зокрема COVID-19. Дозволяє кластеризувати цикл зверху вниз або методи знизу вгору. В даному випадку стратегії кластеризації класифікуються на агломеративні та розділові. Результати ієрархічної кластеризації часто подається у формі дендрограми (деревоподібній формі).

– Кластеризація на основі щільності використовується для формування кластерів, котрі пов'язані між собою на основі зібраних одна ближче до одної точок. В процесі кластеризації відомостей щодо COVID-16 на основі щільності виявляється аномалія точок, проміжки яких є в безпосередньому наборі даних. Цей алгоритм кластеризації найчастіше використовується для опрацювання відомостей щодо COVID-19 завдяки високій продуктивності.

2.1.10 Асоціація відомостей щодо COVID-19

Асоціація – одна з найвідоміших технік у видобуванні даних, в тому числі при видобуванні даних щодо COVID-19. При використанні методів асоціації для видобування відомостей щодо COVID-19 закономірність встановлюється у відповідній операції на основі кореляції між елементами [37]. При використанні методів асоціації метою є встановлення взаємозв'язків у наборах даних щодо COVID-19, на основі класифікування існуючих наборів даних з метою отримання кращого результату. Застосовується у випадках, де присутня потреба забезпечення кращої точності. Присутні дві категорії завдань асоціації – це класифікація майнінгу та майнінг правил асоціації. Поєднання цих двох категорій не потребує ніяких атрибутів для формування та відкриття правил безконтрольного навчання.

2.1.11 Прихований розподіл Діріхле

Прихований розподіл Діріхле (LDA) є одним із найбільш часто використовуваних тематичних методів, де темою є розподіл слів щодо COVID-19, що зустрічаються спільно [38]. Однак теми, вивчені LDA, недостатньо конкретні, щоб відповідати симптомам депресії від COVID-19 та судженням людини [39]. Безконтрольний метод може працювати з некласифікованим текстом, але часто спричиняє перекриття тем [40]. Пізніше метод LDA був розширений, використовуючи терміни, сильно пов'язані з симптомами депресії від COVID-19, як джерела тематичних кластерів і керував моделлю для об'єднання семантично пов'язаних термінів в той самий кластер [41]. Однак такий підхід виявляє лише наявність, тривалість та частоту симптомів стресу, ігноруючи просторовий контекст або фактори навколишнього середовища, які є важливими для моделювання психічного стресу, пов'язаного з COVID-19. Для ідентифікації тексту, пов'язаного з PHQ, та не пов'язаного з ним тексту, був використаний індекс аналізу настроїв, сформований «Python TextBlob» [42], який лише обчислює середню полярність та суб'єктивність кожного слова в даному тексті за допомогою постійного словника [43]. Робота, заснована на імовірнісній генеративній моделі LDA, має обмеження, пов'язані з інтерпретацією факторів введення великих розмірів відомостей щодо COVID-19, що ускладнює узагальнення генеративних моделей без детальних та реалістичних припущень щодо процесу формування даних [44].

У статті [45] запропоновано алгоритм CorExQ9, який інтегрує алгоритм навчання з кореляційного пояснення (CorEx) та клінічний лексикон PHQ для виявлення симптомів стресу, пов'язаних з COVID-19, у просторово-часовому масштабі. Алгоритм використано для дослідження симптомів стресу, спричинених розвитком та поширенням COVID-19 у людей в різних географічних регіонах. Оскільки дані щодо COVID-19 у Twitter – це великі

вхідні дані людей з різними термінами, що використовуються для вираження емоцій, було використано алгоритм CorEx, метод, призначений для обходу обмежень впровадження LDA та мінімізації втручання людини [46]. Після цього розроблено нечітку модель оцінювання точності для візуалізації невизначеності аналітичних результатів на карті поширення COVID-19.

2.2 Збирання та попереднє опрацювання соціальних відомостей щодо COVID-19

Відомості щодо COVID-19, збираються засобами Twitter API. Зібрані дані містять порядка 80 мільйонів твітів за квартал, що приблизно займає 70Гб, що створює значні обчислювальні завдання для традиційного обчислювального ГІС-середовища. Для вирішення цієї проблеми можна використовувати обчислювальне середовище Jupyter, розгорнуте на високопродуктивному комп'ютері Texas A&M. Відфільтровані дані щодо COVID-19 суміжних об'єктів (хештегів, тенденцій та новин) зібрані з використанням Twitter. Потім відбувається видалення невідповідної інформації (твітів, що не стосуються COVID-19, пунктуації, відсутніх даних, брудні коди, URL-адреси, імена користувачів, хештеги, номери та умови запитів) із відфільтрованих твітів. Деякі коригування та нормалізації (наприклад, рівномірний нижній регістр, векторизовані твіти, стандартизація твітів із часовими мітками) також були зроблені для того, щоб задовольнити загальні вимоги моделей машинного навчання. Однак слова зупинки слід видаляти пізніше під час застосування алгоритму для узгодження фразування твітів з лексиконом. Після цього твіти слід токенізувати за допомогою TweetTokenizer Natural Language Toolkit (NLTK) [47]. Слід замінити повторювані послідовності символів, використовуючи значення довжини три для будь-яких послідовностей довжиною три або більше символів, оскільки більшість користувачів часто розширює слова або додає зайві символи, щоб

виразити посилення почуттів. Твіти з точним геопросторовим тегом та міткою часу слід зіставляти з відповідними геолокаційними даними з використанням методів зворотного геокодування [48], [49]. Інші твіти (наприклад, без геотегів, але такі містять інформацію про місцезнаходження, визначену користувачем у профілі), слід геокодувати, використовуючи нечіткий метод пошуку та набір даних псевдонімів населеного пункту [50]. Слід виключили твіти про COVID-19, у яких не було геотегів чи інформації про місцезнаходження, яку визначає користувач. Одним із ключових нововведень є відображення симптомів стресу, спричинених COVID-19, у часовому масштабі. У цьому випадку слід встановили тижневу часову шкалу, тому кількість твітів, зібраних у кожній локації, може бути достатньою для аналізу відомостей про COVID-19.

2.3 Особливості інтеграції даних щодо COVID-19

Обширна множина наукових праць присвячена дослідженню процесів інтеграції великих обсягів даних щодо COVID-19 зібраних з різнотипових джерел [51], [52], [53], [54], [55]. При цьому окрема категорія наукових праць присв'ячена проблемам захисту конфіденційних даних [56], [57], [58], [59], [60]. Зосередимось детальніше на аспектах цих процесів.

2.3.1 Шифрування COVID-19 даних

Шифрування персональних відомостей щодо COVID-19 – це класичний та ефективний, досліджуваний впродовж тривалого періоду часу, спосіб захисту конфіденційних користувацьких даних. У роботі [61] Пенг (Peng) та співавтори подають метод безпечного пошуку на основі сортування за кількома ключовими словами, який застосовує симетричний спосіб шифрування за допомогою відкритого ключа, щоб дозволити користувачеві

зробити безпечний пошук інформації в зашифрованому наборі даних щодо COVID-19 на основі декількох ключових слів. Перевагою методу є реалізація сервісного захисту хмарних обчислень з обмеженими ресурсами. Недоліком є недостатньо висока обчислювальна ефективність. Крім того, присутні ключові ризики розкриття інформації. У роботі [62] Дай та співавтори запровадили своєрідний метод шифрування овальної кривої для реалізації безпечного використання даних щодо COVID-19 і довели, що метод, заснований на шифруванні овальної кривої, перевершує традиційний метод на основі FP. Перевага полягає у відносно високих показниках захисту даних. Однак розглянуто лише простий пошук за ключовими словами на основі булевих функцій, який певною мірою звужує область застосування методу. У роботі [63] Пхаонг та інші використовували векторну космічну модель та гомоморфну техніку шифрування для реалізації ранжирування зашифрованих даних, а також пошуку даних щодо COVID-19 та файлів із кількома ключовими словами. Перевага методу у гарантіях якісного захисту даних. Недоліком є часто дуже високі додаткові витрати обчислювальні, часові та зв'язку. Що стосується проблеми сортування даних з декількома ключовими словами, автори [64] запропонували метод пошуку даних щодо COVID-19 на основі гомоморфного шифрування. При цьому кожен елемент даних щодо COVID-19, готовий до пошуку, є гомоморфним та зашифрованим. Запропонований метод може задовільнити більшість вимог щодо безпечної обробки даних, однак не може забезпечити можливості нечіткого отримання відомостей.

2.3.2 Забезпечення конфіденційності відомостей щодо COVID-19

Покращений метод спільної фільтрації на основі конфіденційності, який називається IPriCF, був поданий у роботі [65] для забезпечення конфіденційності користувачів, що беруть участь у процесі спільної

інтеграції даних щодо COVID-19. Завдяки розділенню персональних даних користувачів та тематичних відомостей щодо COVID-19, IPriCF може ефективно усунути порушення, спричинені шумами на основі різних рівнів конфіденційності. Цей метод може збалансувати конфіденційність даних користувачів та точність рекомендованого списку. Для аналізу різнорідних даних та оптимізації процесів надання послуг в [66] побудована матриця з ознаками елементів зацікавлених сторін. Автори декларують гарантії збереження конфіденційності даних, зберігаючи прийнятну втрату точності прогнозування. Метод факторизації з різною приватною матрицею під назвою DPMF був запропонований у роботі [67]. При цьому була використана техніка факторизації матриць для перетворення чутливих даних користувачів у потенційні вектори низьких розмірів; тоді як диференційована техніка конфіденційності використовувалась для того, щоб приховати цільові функції об'єкта. Однак, із зростанням розмірів, точність прогнозування відповідно зменшується. Автори в роботі [68] вдосконалили модель TrustSVD, запровадивши диференційовану конфіденційність і пропонують нову модель під назвою DPTrustSVD. Нова модель може ефективно досягти компромісу між конфіденційністю даних, розрідженістю даних та доступністю даних. Інша подібна робота [69], в якій автори об'єднали диференціювання приватності та код Хафмана для формування методу публікації сегментів розташування з урахуванням конфіденційності. В [70] автори об'єднали диференціювання приватності, мережу Байєса та теорію ентропії, щоб сформувати метод захисту даних великих розмірів.

2.3.3 Анонімізація даних

Анонімізація є ефективним способом захисту конфіденційних даних користувачів щодо COVID-19 під час аналізу та майнінгу великих даних [71]. Шляхом приховування певної конфіденційної інформації (наприклад, імені,

посвідчення особи), що міститься в даних, анонімізація дозволить безпечно опублікувати решту даних щодо COVID-19. Тобто дані після анонімізації стають безпечно доступні для широкого загалу, для досягнення компромісу між конфіденційністю та доступністю даних щодо COVID-19 [72]. Рішення щодо К-анонімності запропоновано в роботі [73], для приховування ключової конфіденційної інформації у процесі прийняття рішень на основі даних щодо COVID-19. У [74] запропоновано метод захисту місцезнаходження користувача на основі анонімності, для приховування реального розташування або позиції користувачів. Незважаючи на те, що розглянуті вище рішення можуть допомогти приховати конфіденційні дані користувачів під час проведення аналізу відомостей щодо COVID-19 та їх використання в застосунках, керованих на основі даних, вони не можуть збалансувати конфіденційність даних та їх використання, оскільки анонімні дані щодо COVID-19 частково втратять певну ключову інформацію.

Внаслідок аналітичного опрацювання наукових джерел [75], [76], [77], [78], [79], [80] помітно, що, хоча пропонується багато рішень щодо питань передачі та моніторингу даних про COVID-19, вони не забезпечують балансування між багатьма суперечливими критеріями, зокрема безпекою даних, їх наявністю, тощо. Враховуючи зазначений недолік присутня потреба у кращих методах злиття даних щодо COVID-19 зібраних із декількох джерел за допомогою інших видів техніки збереження конфіденційності.

2.4 Інструменти для видобування відомостей щодо COVID-19

Інструменти видобування даних допомагають аналізувати великі за обсягом набори та колекції складних даних щодо COVID-19 на основі виявлених атрибутів набору даних, які користувачі вказують при визначенні тенденцій їх появи. Програмне забезпечення може використовуватися для діагностики симптомів, прогнозування появи та лікування COVID-19 для

видобування нових знань та прийняття рішень [81]. Вибір відповідного програмного забезпечення для вирішення конкретної проблеми в умовах пандемії COVID-19 складний через наявність різнотипових програмних засобів [82]. Найпоширенішими інструментами видобутку даних щодо COVID-19 є наступні.

WEKA [83] (середовище Waikato для аналізу знань): WEKA – це програмний інструмент, який використовується в процесах видобування даних та може ефективно використовуватись для опрацювання відомостей щодо COVID-19. Програмне забезпечення WEKA розробляється з використанням засобів програмування Java і може працювати під управління різних операційних систем. WEKA доповнює декілька процесів видобування даних. Програмне забезпечення може безпосередньо посилатися на дані або бути викликане з програмного Java-коду. Воно використовує графічний інтерфейс користувача (GUI) для контролю продуктивності та підключення додаткових функцій. KEEL [84] використовується для видобування знань на основі еволюційного навчання. Для видобування шаблонів з наборів даних щодо COVID-19 KEEL використовує методи кластеризації, регресійний аналіз та методи класифікації. Програмний пакет KEEL постачається як програмне забезпечення з відкритим кодом. Проте вихідний програмний код пакету KEEL приховано. Інструменти видобування даних з пакету KEEL можна використовувати для повного аналітичного опрацювання колекцій та наборів даних щодо COVID-19.

R [85] – це програмний продукт з відкритим кодом, який використовується для здійснення обчислень та статистичного аналізу відомостей щодо COVID-19. Програмне забезпечення R приносить велику користь для світу досліджень та розробок та для сектору охорони здоров'я.

KNIME [86] використовується для моделювання та аналітичного опрацювання відомостей щодо COVID-19 та поширюється як програмне забезпечення з відкритим кодом. Програмно-аналітичний пакет KNIME

містить функції машинного навчання та видобування даних, котрі можна ефективно використовувати для аналітичного опрацювання відомостей щодо COVID-19. Зокрема KNIME застосовується в клінічних дослідженнях, виявленні та оцінці COVID-19. Програмний пакет KNIME може створювати спеціалізовані робочі процеси, які можна подавати та записувати у різних форматах. RAPIDMINER [87] окрім ефективного використання для аналітичного опрацювання відомостей щодо COVID-19 може бути використано для організації процесів видобування даних у різних сферах. Зокрема для потреб малого та великого бізнесу, у фінансовій галузі, банківській та страховій справі, науці та освіті. RAPIDMINER – це програмно-алгоритмічний комплекс з відкритим кодом, яка використовується в численних сферах людських зусиль.

Програмно-алгоритмічний комплекс Orange [88] поширюється у формі програмного забезпечення з відкритим кодом і представлене у формі двох частин: фронтальної та «бакенд». Для реалізації функціоналу фронтальної частини використовуються парадигми візуального програмування, а для «бакенд» використовуються бібліотеки python. Програмно-алгоритмічний комплекс Orange розроблено з використанням середовища програмування C++ та Python. Програмно-алгоритмічний комплекс Orange використовується у наукових цілях для проведення генетичних досліджень та у медичній галузі для дослідження COVID-19.

2.5 Висновок до другого розділу

В другому розділі кваліфікаційної роботи описано Методи видобування відомостей щодо COVID-19. Розглянуто збирання та попереднє опрацювання соціальних відомостей щодо COVID-19. Подано особливості інтеграції даних щодо COVID-19. Досліджено інструменти для видобування відомостей щодо COVID-19.

3 ПРАКТИКИ ВИКОРИСТАННЯ АНАЛІТИЧНИХ ІНСТРУМЕНТІВ ДЛЯ ОПРАЦЮВАННЯ ВІДОМОСТЕЙ ЩОДО COVID-19

3.1 Злиття та видобування відомостей щодо COVID-19 отриманих з ІоН-джерел

На рисунку 3.1 подано структуру злиття відомостей щодо COVID-19, зокрема медичних записів лікарів, медичних сестер, пацієнтів котрі частково розміщені на хмарних платформах ХП1 та ХП2 відповідно.

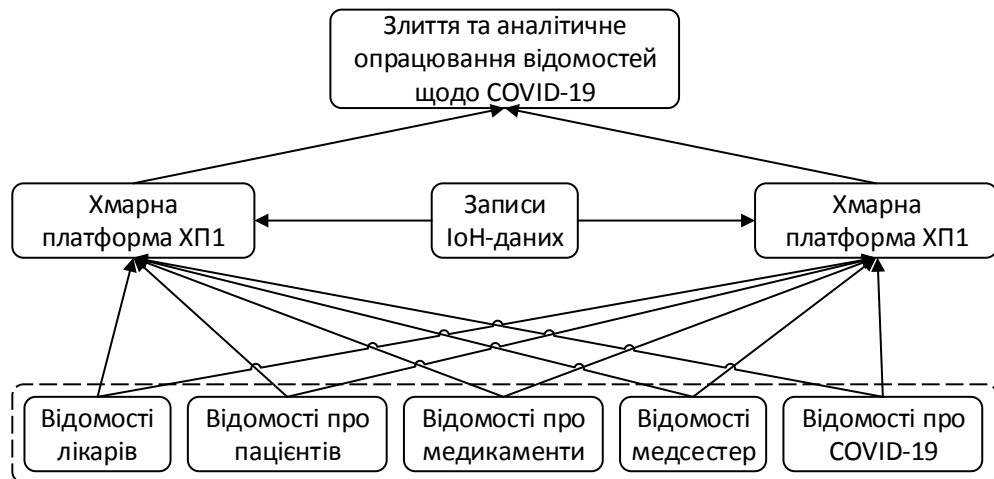


Рисунок 3.1 – Злиття відомостей щодо COVID-19 отриманих з використанням ІоН-джерел

Щоб здійснити всебічне аналітичне опрацювання та видобування цінної інформації щодо COVID-19 з ІоН-даних, розподілених на платформах ХП1 та ХП2, потрібно інтегрувати та об'єднати дані з кількох джерел для єдиного аналізу даних та прийняття нових науково-обґрунтованих рішень в галузі охорони здоров'я.

Однак у вищезазначеному процесі злиття та аналізу відомостей щодо COVID-19 на основі ІоН-даних часто виникають проблеми щодо конфіденційності, оскільки історичні записи ІоН-даних часто містять конфіденційну інформацію пацієнтів. Щоб стимулювати хмарні

інформаційно-технологічні платформи ХП1 та ХП2 надавати свої записи ІоН-даних та мінімізувати загрозу розкриття конфіденційності інфікованих COVID-19 пацієнтів, необхідно розробити ефективні методи злиття даних, не розкриваючи конфіденційність.

Тому в даному розділі кваліфікаційної роботи досліджено методи злиття ІоН-даних отриманих з декількох джерел без загрози конфіденційності. Для кращого опису деталей методів злиття відомостей щодо COVID-19 без розкриття конфіденційної інформації, проведемо узагальнення використаних символів та наведемо відповідні позначення в таблиці 3.1.

Таблиця 3.1 – Специфікації символів та позначень

Позначення	Опис
R_1, \dots, R_n	Елемент ІоН-даних щодо COVID-19
q_1, \dots, q_m	Критерій охорони здоров'я
f_1, \dots, f_a	Хеш-функція
T_1, \dots, T_b	Хеш-таблиця
cp_1, \dots, cp_h	Розподілена хмарна платформа
v_1, \dots, v_m	М розмірність кожної хеш-функції
$h_1(R_x), \dots, h_a(R_x)$	Хеш-значення R_x базоване на f_1, \dots, f_a
$H_1(R_x), \dots, H_b(R_x)$	Індекси R_x в хеш таблицях T_1, \dots, T_b

Основну процедуру методу злиття та видобування відомостей щодо COVID-19 можна узагальнити наступними кроками. По-перше, важливі ІоН-дані проектуються на основі LSH-функцій. По-друге, відповідно до кожного запису ІоН-даних та відповідних йому хеш-значень, отриманих після проєкції хешу, створюється набір хеш-таблиць без конфіденційних відомостей пацієнтів, інфікованих COVID-19. На третьому етапі, згідно з

отриманими хеш-таблицями, відбувається аналогічний пошук і видобування ІоН-даних щодо COVID-19 (див. рисунок 3.2).

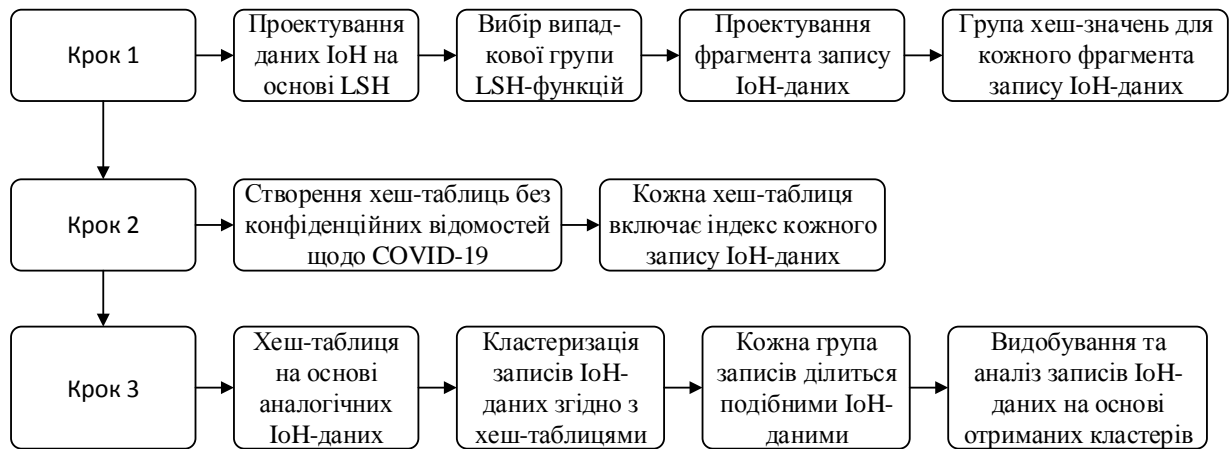


Рисунок 3.2 – Етапи злиття та видобування відомостей щодо COVID-19

Проведемо детальніший опис етапів злиття та аналітичного опрацювання відомостей щодо COVID-19.

1-й крок включає проектування структури ІоН-даних щодо COVID-19 на основі LSH. Для цього використовується R_{i,q_j} для позначення розмірності q_j ($j = 1, 2, \dots, m$) історичних записів ІоН-даних від R_i ($i = 1, 2, \dots, n$) пацієнта. Оскільки R_{i,q_j} чутливий до відомостей пацієнта, потрібно забезпечити приватність інформації R_{i,q_j} , коли R_{i,q_j} публікується для широкого загалу. На цьому кроці для досягнення мети використовується LSH-стратегія.

Для R_i ($i = 1, 2, \dots, n$) присутні m критеріїв q_1, \dots, q_m . Корисна інформація, що відповідає R_i , може бути зображеним символом $\vec{R}_i = (R_{i,q_1}, \dots, R_{i,q_m})$. Коли випускаються \vec{R}_i іншим, потрібно спочатку зробити LSH-проекцію. Створюється новий вектор $V = (v_1, \dots, v_m)$, де v_j ($j = 1, 2, \dots, m$) – випадково згенерованих значень з домену $[-1, 1]$. Таким чином, створюється LSH-функція f , як у рівнянні (3.1). Згодом можна буде отримати $f(\vec{R}_i)$, який може бути позитивним або негативним. Далі формується наступне відображення відповідно до формули (3.2).

$$\begin{aligned}
 f(\vec{R}_i) &= \vec{R}_i \cdot \vec{V} \\
 &= (R_i \cdot q_1, \dots, R_i \cdot q_m) \cdot (v_1, \dots, v_m) \\
 &= R_i \cdot q_1 \cdot v_1 + \dots + R_i \cdot q_m \cdot v_m
 \end{aligned}
 \tag{3.1}$$

Таким чином, шляхом відображення відбувається перетворення $f(\vec{R}_i)$ в $h(R_i)$ у формі двійкового значення 0 або 1.

$$h(R_i) = \begin{cases} 1, & \text{if } f(\vec{R}_i) > 0 \\ 0, & \text{if } f(\vec{R}_i) \leq 0 \end{cases}
 \tag{3.2}$$

Конкретна процедура виконує перетворення вхідних сутностей у вихідні. Вхідні сутності:

- Історичні записи ІоН-даних щодо COVID-19 – R_1, \dots, R_n .
- Вимірювання якості ІоН-даних – q_1, \dots, q_m .

Вихідні сутності:

- $h(R_i)$ – логічне значення R_i після відображення.

Відповідний програмний код подано в лістингу 3.1.

Лістинг 3.1 – Програмний код процедури перетворення даних

```

1: for j = 1, ..., m do
2: vj = random[-1,1]
3: end for
4: for i = 1, ..., n do
5:   sum = 0
6: for j = 1, ..., m do
7:   sum + = Ri.qj* vj
8:   end for
9:   if sum > 0
10: then → f(Ri) = 1
11: else f(Ri) = 0
12:   end if
13: return f(Ri)
14: end for

```

2-й крок включає створення хеш-таблиць без відомостей щодо конфіденційності. При цьому виведене на першому етапі $f(\vec{R}_i)$ може розглядатися як хеш-значення R_i . Однак одного процесу прогнозування недостатньо для перетворення R_i в індекс без конфіденційності. Враховуючи слід неодноразово повторювати поданий вище алгоритм над проекціями f_1, \dots, f_a . Після чого буде отримано адиментальний хеш-вектор $H(R_i)$, поданий у рівнянні (3.3).

$$H(R_i) = (h_1(R_i), \dots, h_a(R_i)) \quad (3)$$

Таким чином, відображення $R_i \rightarrow H(R_i)$ ($i = 1, 2, \dots, n$) складають хеш-таблицю, що позначається як «Т». Іншими словами, через «Т» записуються значення індексу R_i . Але, навпаки, враховуючи значення індексу R_i , неможливо зробити висновок про реальне значення R_i . Таким чином забезпечується конфіденційність інформаційних записів про інфікованих COVID-19 пацієнтів, що містяться в множині R_i .

Однак одна хеш-таблиця «Т» може не завжди точно відобразити реальний індекс кожного запису ІоН-даних.

Вхідні сутності:

- $h(R_1), \dots, h(R_n)$ – булеві значення записів ІоН-даних.
- f_1, \dots, f_a – LSH- функції.

Вихідні сутності:

- T_1, \dots, T_b – хеш-таблиці.

Враховуючи це обмеження, слід декілька разів повторювати процес створення хеш таблиць «Т». В результаті буде отримано b таблиць – T_1, \dots, T_b . Цей крок можна формалізувати псевдокодом поданим в лістингу 3.2.

Третій крок-3: аналіз ІоН-даних на основі хеш-таблиць. На другому кроці генеруються таблиці b – T_1, \dots, T_b . В кожній таблиці міститься набір відповідних пар $R_i \rightarrow H(R_i)$ ($i = 1, 2, \dots, n$).

Лістинг 3.2 – Псевдо-код хешування

```

1: for x = 1, ..., a do
2: repeat Algorithm 1 based on  $f_x$ 
3: end for
4: for i = 1, ..., n do
5:  $H(R_i) = (h_1(R_i), \dots, h_a(R_i))$ 
6: put  $R_i \rightarrow H(R_i)$  into T
7: end for
8: return T
9: repeat lines 1-8 b times

```

Крім того, $H(R_i)$ вважається непридатним для використання з приблизним індексом R_i . Згідно з теорією чутливого хешування, записи ІоН-даних з однаковим індексом були б приблизно однаковими. Як результат, якщо два записи R_1 і R_2 мають однаковий індекс, то R_1 і R_2 , ймовірно, подібні записи. Таким чином, можна видобувати потенційні схожі записи ІоН-даних щодо COVID-19, перевіряючи відповідні значення їх індексів без зайвих ризиків конфіденційності. Однак для двох записів ІоН-даних щодо COVID-19 R_1 і R_2 $H(R_1) = H(R_2)$ є досить жорсткою умовою обмеження, оскільки кожне розмірне значення $H(R_1)$ має бути точно рівним $H(R_1)$. Така жорстка умова обмеження здатна повертати порожній результат пошуку аналогічних записів ІоН-даних. Враховуючи цей недолік, для зменшення вищезазначеного жорсткого стану слід створювати більше хеш-таблиць замість однієї. При цьому конкретно враховуються таблиці b , створені на другому кроці, тобто T_1, \dots, T_b . Якщо $H(R_1) = H(R_2)$ виконується в будь-якому T_y ($y = 1, 2, \dots, b$), то робиться висновок, що R_1 і R_2 , ймовірно, подібні записи ІоН-даних щодо COVID-19. Таким чином пом'якшується аналогічний стан пошуку відповідних записів ІоН-даних. Для конкретного запису ІоН-даних R_x можна шукати подібний набір записів $\text{Sim_Set}(R_x)$ за допомогою поданої вище послідовності дій. Деталі цього етапу подані в лістингу 3.3. Після чого в якості кінцевого результату повертається $\text{Sim_Set}(R_x)$.

Вхідні сутності:

– T_1, \dots, T_b – b хеш-таблиці;

- R_1, \dots, R_n – історичні записи ІоН-даних щодо COVID-19.
- R_x – цільові записи ІоН-даних, подібні до потрібних.

Вихідні сутності:

- $\text{Sim_Set}(R_x)$ – множина подібних записів ІоН-даних щодо R_x .

Лістинг 3.3 – Програмний код для пошуку подібних сутностей

```

1: Sim_Set (Rx) = ∅
2: for y = 1 to b do
3:   for i = 1, ..., n do
4:     if H(Ri) = H(Rx)
5:     then put Ri into Sim_Set (Rx)
6:     end if
7:   end for
8: end for
9: return Sim Set (Rx)

```

Щоб підтвердити ефективність розробленого програмно-алгоритмічного рішення щодо синтезу та видобування даних (скорочено PDFM), проведено ряд експериментів та порівняно з існуючими підходами.

Використовуються загальнодоступні дані [89], опубліковані для моделювання. У наборі даних кожна пара useritem-qos приймається як пара пацієнт-критерій-значення у сценаріях синтезу ІоН-даних щодо COVID-19 із кількома джерелами. 90% записів набору даних використовуються для навчання параметрів моделі щодо злиття та видобування відомостей про COVID-19, тоді як 10% записів даних використовується для тестування та перевірки. Щоб показати конкурентні переваги PDFM, для порівняння з PDFM використовуються методи UCF (базовий рівень) та ICF. Порівнянні показники включають точність прогнозування відсутніх даних (Середня абсолютна помилка) та обчислювальний час. Програмно-апаратна конфігурація сформована на базі процесора з тактовою частотою 2,80 ГГц, пам'ять 8,0 ГБ, операційною системою Windows 8.1 та JAVA 8. Кожен експеримент запускався не менше п'ятидесяти разів та обчислювалось їх середнє значення.

3.2 Використання штучного інтелекту для опрацювання відомостей щодо COVID-19

Практично до 2018 року спостерігався невпинний прогрес галузі ШІ котрий знайшов широке застосування в сучасній науці та медицині [90]. ШІ допомагає медичним працівникам та керівництву усіх рангів у процесах прийняття рішень. Керуючись інноваційними клінічними дослідженнями, основні стратегії ШІ допомагають відкрити суттєві, важливі та значимі дані, приховані у великих за обсягом наборах медичних відомостей [91]. Підтвержені випадки COVID-19 загрожують надзвичайно інтенсивними розладами дихання. За співвідношенням підтвердженої випадків смертності COVID-19 перевершив ГРВІ [92]. COVID-19 поширюється серед населення використовуючи порівняльну механіку дії вірусу для проникнення та забруднення клітин організму [93]. Тому використання системи на основі ШІ для боротьби з COVID-19 є перспективним напрямком досліджень, наприклад, з метою передбачення областей наступного спалаху [94]. Такий програмно-алгоритмічний комплекс успішно прагнула зробити канадська організація Blue Dot. Різні види застосування систем на основі ШІ, які з'явилися у світі у світлі останньої пандемії COVID-19, включають Benevolent AI та Імперський коледж Лондона, які повідомляють, що ліки, підтвержені проти ревматоїдного запалення суглобів, барицитинів, можуть бути ефективними проти інфекції COVID-19. Тим часом гонконзька компанія Insilico Medicine заявила, що за розрахунками ШІ спроектовано шість нових молекулярних структур, які можуть припинити вірусну реплікацію [95]. Швидке вдосконалення медичної інформації супроводжується експоненційними темпами зростання обсягів медичних даних, що ускладнює ризики в галузі медичних виробів для сучасних лікарняних інформаційних систем [96]. Галузь охорони здоров'я генерує широкий перелік різнотипових наборів даних щодо діагностування, записів про пацієнтів, лікування, тощо.

Через обмеженість розроблених на даний час процедур опрацювання даних точність певних звітів впливає на організацію системи охорони здоров'я [97]. Використання ШІ може покращити результати використання медичних інформаційних систем, які збирають великі обсяги фактичної інформації про пацієнтів [98]. Великі за обсягом набори даних в секторі охорони здоров'я вже ефективно використовуються в окремих медичних закладах та установах. Зокрема у рентгенології за допомогою алгоритмів глибокого навчання відбувається діагностування захворювань [99]. Візуалізація великих даних відіграє важливу роль не тільки для управління медичними закладами, але й в процесах утилізації медичних відходів [100].

Перш ніж медичні засоби, сформовані на основі ШІ, можуть бути задіяні в програмах запобігання та протидії COVID-19, вони повинні пройти процеси навчання на основі відомостей отриманих в результаті медичних експериментів. Це можуть бути початкові набори даних, наприклад, щодо скринінгової діагностики, аналізу вірусів, лікувальних процедур, тощо. Ця клінічна інформація регулярно оновлюється та не обмежується типовими демографічними чи клінічними записами, електронними звітами клінічних приладів, фізичними оцінками, клінічними дослідженнями та зображеннями [101]. На рисунку 3.2 проілюстровано обсяги аналітичного опрацювання даних в медичній галузі з використанням засобів ШІ, включаючи аналіз зображень, тести спадковості та електродіагностику [102].

Автори [103] досліджують рентгенограми засобами систем ШІ для дослідження зображень. У роботі [104] розглядається можливість використання незвичної спадкової артикуляції у проміжках некодуючих РНК для аналізу злоякісних утворень шлунка. Автори [105] розробили електродіагностичні засоби для емоційно підтримуючої мережі що використовується при травмах мозку. Оцінювання фізичного стану та результати експериментальних медичних центрів розглядаються як частина інших важливих інформаційних областей для систем ШІ (Див. рисунок 3.3).

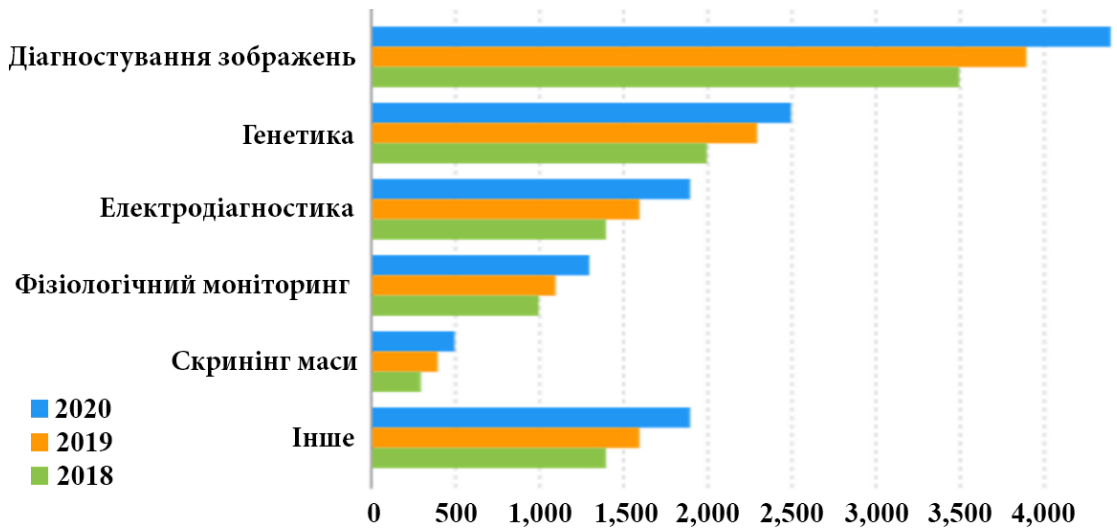


Рисунок 3.3 – Типи медичних відомостей, що досліджуються з використанням систем ІІІ

Прикладом застосувань на основі систем ІІІ є електронна медична карта (EMR). Але автори [106] використовували досягнення ІІІ для видалення основних фенотипових моментів зі звітів, щоб підвищити точність знаходження вроджених порушень. У [107] описано особливості використання ІІІ для діагностування та спостереження за перебігом COVID-19. Оскільки постійно зростає кількість досліджуваних зразків, ці обчислювальні методи покращують точність діагностування. У роботі [107] в відзначено, що передача медичних клінічних симптомів від людини до людини складає 41% справ у розслідуванні ланцюжків інфікування пацієнтів лікарні в м. Ухань. Понад 1000 практикуючих лікарів швидкої медичної допомоги підтверджуються забрудненими внаслідок контакту з інфікованими COVID-19 пацієнтами. При КТ-перевірці легенів ІІІ використовується для ідентифікації пошкоджень внаслідок пневмонії викликаной COVID-19. Для цього вимірюють їх об'єм, форму та товщину. Після чого система на основі ІІІ аналізує зміну численних осередків ушкоджень легенів на знімках та допомагає для встановлення діагнозу. Аналіз результатів КТ вручну може зайняти більше 15 хвилин, тоді як ІІІ опрацює зображення не більше ніж за 10 секунд.

3.3 Методи навчання систем ШІ при діагностуванні COVID-19

Суттєвим зауваженням клінічних спеціалістів при використанні систем комп'ютерного зору на основі ШІ є чіткість та точність діагностування. При цьому спостерігалась занадто поміркована інтерпретація від моменту дослідження до практичних репетицій, незалежно від того, чи відбувалася робота з галуззю соціального страхування як аргумент для пришвидшення тлумачення зображень фахівцями. Як відзначають експерти, збільшення фінансування досліджень, дозволяє успішно вдосконалювати ШІ та гарантувати його якість [108]. Проте використання систем ШІ спричиняє проблеми з персоналом медичних закладів. Зокрема в лікарів зустрічається синдром «Дескілінг» (англ. Deskilling) – це втрата здібностей лікаря, який не має найменшого уявлення про те, як виконувати раніше виконувані завдання, які виконував раніше, оскільки перестав їх виконувати, щоб підтримати машини.

3.3.1 Контрольоване навчання

Контрольоване навчання – це один з загальноприйнятих та найбільш використовуваних методів застосування ШІ у галузі охорони здоров'я. Цей метод навчання використовує наявну інформацію для формування максимально-точних прогнозів і засвоює кореляцію між бажаними результатами та відповідними вхідними наборами. При цьому навчальні інформаційні набори аналізуються на основі навчальної стратегії для розпізнавання сутностей в залежності від близьких характеристик. На даний час використовуються різні способи передбачення результату [109].

На сьогодні в медицині поширені інформаційні системи з використанням елементів ШІ на основі методу опорних векторів, нейромерех та дерев рішень. При встановленні оцінок, як правило, межа описується в

медичній інформаційній системі. Це допомагає забезпечити найкращий взаємозв'язку даних та результатів. Окремі навчені в такий спосіб моделі забезпечують точність класифікації медичних даних до 90%. Контрольоване навчання допоможе сформувати черговий етап для програмно-алгоритмічного спостереження та прогнозування COVID-19. Також можуть бути створені нейромережні інформаційно-технологічні системи для виокремлення візуальних аспектів цього захворювання. Це допоможе у перевірці та лікуванні інфікованих COVID-19 людей. Для навчання нейромереж можна подавати найновіші відомості та рентенограми пацієнтів щоб якомога оперативніше реагувати на запити, пов'язані з COVID-19. Контрольований підхід до навчання може стимулювати підготовку нейромережних інформаційно-технологічних систем [110].

Метод розпізнавання рентгенівських зображень «Xception», сформований на основі глибинного підходу CNN з використанням зворотніх розпізнавальних шарів, був запропонований у [111]. Запропонована нейронна мережа складається з двох шарів, розмежованих прихованими шарами згортки, що діляться по глибині, чотирма шарами згортки та повністю пов'язані шаром зворотнього зв'язку. У [112] цей метод був використаний при класифікації положення ліжка пацієнта з використанням різних давачів тиску вмонтованих в ліжку. Завдяки своїм можливостям та високоефективним результатам, зазначений метод може служити корисним інструментом у боротьбі з COVID-19. На рисунку 3.4 подано метод контрольованого навчання. При цьому позначено отримані від пацієнта атрибути навчання та тестування (X), прогнозний результат (Y). Навчання відбувається за допомогою набору підготовлених даних з міткою (X), яке дає прогнозовані значення з набору (Y). Функціональність нейромережі буде формуватися відповідно до цього набору даних. Після навчання ця функціональність буде використана для класифікації нових відомостей отриманих зі значень із підозрою щодо COVID-19.

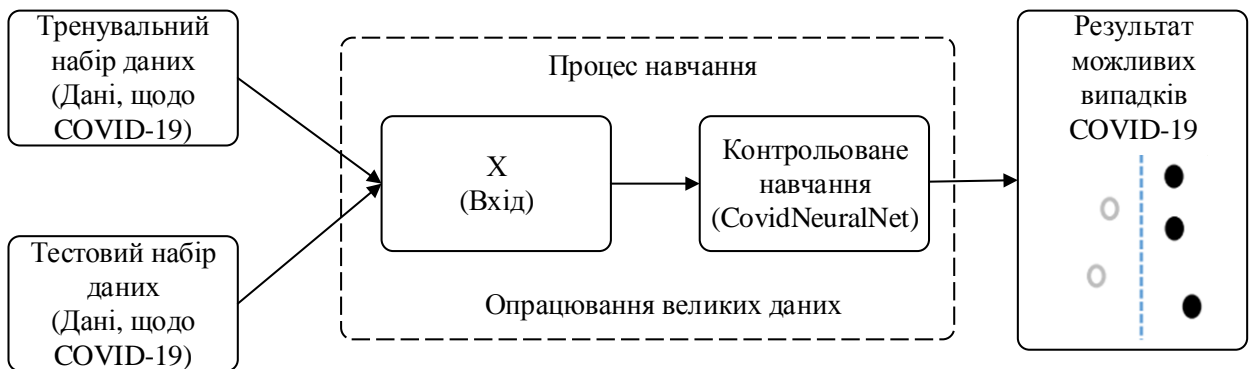


Рисунок 3.4 – Контрольоване навчання нейронної мережі для класифікації відомостей щодо COVID-19

3.3.2 Неконтрольоване навчання

Замість розглянутої в попередньому параграфі методики навчання, яка використовує перевірені дані, методика неконтрольованого навчання не використовує керуючих сигналів та імен. Цей прийом зазвичай використовується для пошуку виокремлених структур даних шляхом розбиття їх на близько-розміщені набори. При цьому в основному використовується для з'ясування властивостей даних, що відрізняють результуючі набори. Це перспективний вид оцінювання наборів даних для досягнення загальної мети системи сформованої на основі ШІ. Проте він не менш ефективний ніж інші методи машинного навчання. К-засоби, та автоматична класифікація є найбільш відомими некерованими методами машинного навчання. Однією з найбільш широко поширених задач цього методу навчання в медичній галузі є визнання різноманіття [113]. Дані, отримані на основі приналежності, починатимуться з еквівалентного розсіювання. Якщо є якісь неточності або підробка, виявлена в наборах даних про приналежність, ці дані будуть визнані винятком і можуть без особливої затримки бути затверджені або відхилені. К-засоби є загальновикористовуваним механізмом групування при аналітичному опрацюванні зображень [114]. Метод потребує від клієнта вказати кількість

кластерів k , які буде сформовано. Аналогічно до [115], метод був використаний для відсікання зашумлених даних від корисних, за рахунок ігнорування невеликих підкластерів в множині даних. Метод може бути застосований для аналітичного опрацювання КТ-зображень та в медичних застосунках для діагностування COVID-19. У [116] автор висунув нову структуру поданої вище моделі навчання для відокремлених груп, яка виокремлює значимі дані з потоку. Потім нейронна мережа використовується для передбачення руху потоку подій. Цей рух використовується для виділення будь-яких знань, прихованих в потокових зображеннях. На рисунку 3.5 подано описаний метод навчання. Атрибути навчання та тестування, отримані від пацієнта з підозрою COVID-19 позначено (X).

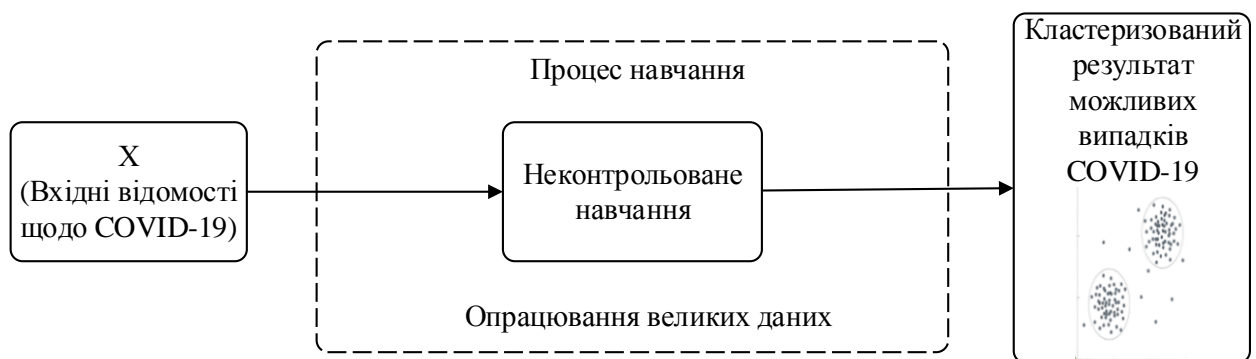


Рисунок 3.5 – Неконтрольоване навчання для кластеризації COVID-19

В описаному вище методі не була сформована мета досягнення точності. Метою підходу виявлення будь-яких зразків, що привертають достатньо уваги для виділення в окремий кластер. При цьому додаткові опрацьовані відомості можуть допомогти у підтвердженні або відхиленні підтвердження виявлених зразків.

3.3.3 Підкріплення навчання

Методологія підкріплення навчання максимально подібна на процес людського дослідження та навчання за допомогою звичайних завдань. Ця

техніка навчання не є жодною з двох вищеописаних. Це різновид методів на основі схрещування, при якому існує альтернативний або єдиний авторитет, який має обґрунтування прийнятної поведінки за певних обставин з кінцевою метою, якою він намагаються збільшити свою винагороду або бали. Зокрема, там, де зафіксовано поточний стан, замовник вибирає найкращі дії щодо пояснення методології, при цьому стан зміниться, і практикуючі сутності отримують нагороду за короткий момент перебування у найкращому стані, або заявляють пізніше, якщо використовується відкладення винагороди. Це змушує практикуючих не обирати дію, виконану в минулому, яка привела до отримання цього призу [117]. Практикуючий навчається оптимальному підходу, який представляє собою відображення стану для переміщення, не переходячи до даних про природу. План гри виявляється лише шляхом проведення клінічних експериментів або неправильного використання та критичного обстеження. Ця модель здатна приймати рішення, використовуючи залучення кооперативів та оціночні відомості. На відміну від звичних стратегій підкріплене навчання під наглядом, які, як правило, залежать від одноразових, ретельних та контрольованих сигналів [118]. Метод використовується для опрацювання послідовних динамічних завдань за допомогою інспекційних, оціночних та відкладених входів. Такі особливі моменти роблять метод підкріпленого навчання обґрунтованим варіантом для реалізації інноваційних механізмів в обширному переліку досліджень в медичній галузі. Зокрема, для діагностування вибору схеми лікування, яка, як правило, описується відкладеною та послідовною стратегією [119]. Ця модель надає можливість фахівцям вивчати життєздатні процедури в послідовних динамічних питаннях шляхом експериментальної комунікації відповідно до сформованих умов [120]. Наприклад, «Процес прийняття рішень Маркова» (MDP) був використаний для вивчення гіпотетичної динаміки в стохастичних умовах, як загальна структура для формалізації питання моделей. MDP може бути реалізований у різних структурах,

спираючись на дані, які можуть бути виокремлені з попереднього етапу. На рисунку 3.6 подано метод підкріпленого навчання.

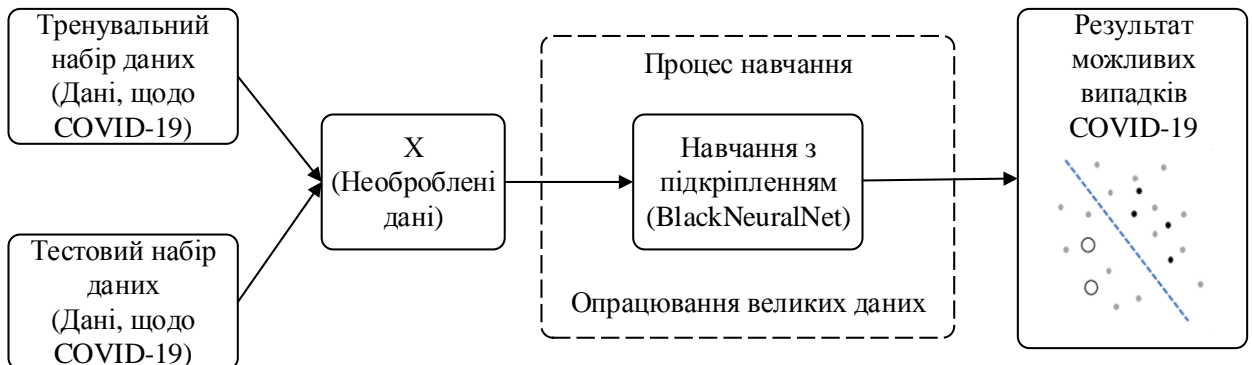


Рисунок 3.6 – Метод навчання з підкріпленням

При цьому символом (X) позначено атрибути навчання та тестування, отримані від інфікованого COVID-19 пацієнта, а результат (Y). В даному випадку потрібен точний прогноз.

3.4 Порівняльна характеристика методів навчання систем ШІ

Порівняльну характеристику методів навчання систем ШІ що використовуються для діагностування та протидії COVID-19 подано в таблиці 3.2.

Таблиця 3.2 – Порівняльна характеристика методів машинного навчання для систем ШІ COVID-19

Метод	Контрольоване навчання	Неконтрольоване навчання	Навчання з підкріпленням
1	2	3	4
Визначення	Видобуває знання, використовуючи марковані дані	Тренується на не маркованих даних.	Практикується взаємодія з оточенням, дії виконуються опираючись на нагороди чи помилки

Продовження таблиці 3.2

1	2	3	4
Тип задач	Класифікація та регресія	Кластеризація та об'єднання.	На основі винагород
Тип даних	Позначені даними	Дані без міток	Відсутність заздалегідь визначених даних
Навчання	Зовнішній нагляд	Ніякого нагляду	Ніякого нагляду
Підхід	Вхідні дані, що позначаються, впливають на відомий результат	Шаблони розуміються і вихід виявляється	Впроваджуються спроби та помилки
Операція	ML	ML	ML
Розвідка	Без розвідки	Без розвідки	Адаптується шляхом розвідки
Стратегія	Алгоритм навчання та залежність від даних	Класифікація та залежність від даних	Вчиться на досвіді

3.5 Методи глибинного навчання та COVID-19

Використання глибинного навчання буде ефективним для виявлення, аналізу, передбачення та розслідування випадків інфікування COVID-19 та сприяння збереженню фінансових наслідків. З моменту спалаху пандемії COVID-19 тривають наукові дебати щодо дослідження та використання засобів ШІ та інших аналітичних засобів [121]. Автор декларує наміри швидкого розгортання та створення великих за обсягом сховищ даних для зберігання відомостей щодо досліджень, медичних заходів та клінічних обстежень. Наслідки пандемії COVID-19 будуть катастрофічними для життя фінансового забезпечення громадян. Тому актуалізуються дослідження методів глибинного навчання, одного з найбільш перспективних інформаційно-технологічних аналітичних інструментів на основі ШІ. Однак на даний час засоби з використанням ШІ ще не проявили свою ефективність

проти COVID-19, оскільки їх використанню ШІ заважають пропуски в колекціях даних, зашумленість даних та присутність аномальних значень. Для усунення перелічених недоліків колекцій відомостей щодо COVID-19 доцільно використати методи глибинного навчання. Збір аналітичної інформації щодо летальних випадків COVID-19 може мати фундаментальне значення для збереження життя громадян та обмеження фінансових втрат. Автори [121] рекомендують використовувати глибоке навчання для просторового відстеження та прогнозування поширення COVID-19. Зокрема після минулої пандемії Вірусу Зіка, була створена потужна нейромережна інформаційна система, для прогнозування її поширення. Розроблені моделі слід повторно підготувати, використовуючи інформацію щодо пандемії COVID-19. Автори [122] подають опис інформаційної системи для швидкого та ефективного аналізу рентгенограм грудної клітки пацієнтів Covid-19 з використанням методів глибинного навчання. Описана інформаційна система використовується для класифікації нерегулярних випадків пневмонії, що допомагає прискоренню аналізу наслідків пандемії. Навчені схеми класифікують нездужання у 87% випадків, порівняно з експертами з надання соціальних послуг. Результати діагностування економлять 92% часу порівняно із залученням для діагностування медичних експертів людей. Автори [122] пропонують два методи аналітичного опрацювання, які можуть бути критично важливими для контролю поширення зростаючої кількості вірусів, серед яких COVID-19. Під час поточного дослідження автори змогли створити програмні засоби на основі методів та стратегії глибинного навчання для аналізу КТ-зображень. Методи глибинного навчання можуть сприяти боротьбі з COVID-19, використовуючись для:

- попередження та запобігання COVID-19;
- прогнозування та відстеження COVID-19;
- діагностики;
- соціального контролю;

– підвищення ефективності лікування.

Програмно-алгоритмічні застосунки на основі методів глибинного навчання використовуються для класифікації, ідентифікації та формування переліків характеристик захворювань на основі опрацювання медичних зображень. Це дозволяє розширити теоретичні знання щодо COVID-19 та сформувати ефективні прогностичні механізми лікування пацієнтів. Зазначені програмно-алгоритмічні застосунки використовуються для аналітичного опрацювання великих за обсягом наборів даних, що дозволить значно зменшити фінансові витрати в медичній галузі. Парадигма Інтернету медичних речей (ІоМТ), яка сформована на основі «розумних» медичних систем, також може ефективно використовуватись для відслідковування важливих проблем здоров'я громадян. Епідемічні хвороби можна відстежувати та уникати шляхом інтеграції великих за обсягом колекцій даних, зібраних із використанням ІоМТ-давачів [122].

3.6 Висновок до третього розділу

В другому розділі кваліфікаційної роботи описано злиття та видобування відомостей щодо COVID-19 отриманих з ІоН-джерел. Досліджено питання використання штучного інтелекту для опрацювання відомостей щодо COVID-19. Проаналізовано методи навчання систем ШІ при діагностуванні COVID-19. Подано порівняльну характеристику методів навчання систем ШІ та описано методи глибинного навчання у боротьбі з COVID-19.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Аналіз нормативних документів, які визначають дії роботодавця, працівник якого захворів на COVID-19

Карантинні заходи пов'язані з COVID-19 тривають вже більше пів року. Статистика щодо інфікування невпинно зростає. При цьому роботодавці стикаються з цілою низкою проблем пов'язаних із COVID-19. Тому актуальним питанням в галузі охорони праці є аналіз нормативних документів, які визначають порядок (план) дій роботодавця, працівник якого захворів на COVID-19 чи є контактною особою.

На даний час нормативних документів щодо COVID-19 немає. Але є бурхлива діяльність уряду та Верховної Ради. Крім того є відповідальність за порушення правил щодо карантину людей, санітарно-гігієнічних, санітарно-протиепідемічних правил і норм, передбачених Законом України «Про захист населення від інфекційних хвороб» від 06.04.2000 р. № 1645-III [123], іншими актами законодавства, а також рішень органів місцевого самоврядування з питань боротьби з інфекційними хворобами. Що загрожує посадовим особам роботодавців адміністративним штрафом від 34000 до 170000 грн (ст. 44-3 КпАП [124]).

Оскільки захворювання на COVID-19 належить до інфекційних захворювань, то слід враховувати Порядок розслідування та обліку нещасних випадків, професійних захворювань та аварій на виробництві, затверджений постановою КМУ від 17.04.2019 р. № 337 [125]. Адже гостре профзахворювання – це захворювання, що виникло після однократного (протягом не більш як однієї робочої зміни) впливу на працівника шкідливих факторів фізичного, біологічного та хімічного характеру (у тому числі інфекційні захворювання). Однозначно визначити, де був контакт та відбулося інфікування COVID-19, не завжди можливо. Саме тому Державна

служба України з питань праці [126] наголошує, що обов'язок провести розслідування випадків гострої респіраторної хвороби COVID-19, відповідно до Порядку № 337, стосується лише лабораторно підтверджених випадків інфікування COVID-19 медичних та інших працівників, що пов'язані з виконанням профобов'язків в умовах підвищеного ризику зараження на COVID-19 (надання меддопомоги хворим на інфекційні хвороби, робота з живими збудниками та в осередках інфекційних хвороб, дезінфекційні заходи тощо).

Якщо працівник звичайного роботодавця хворів на COVID-19, він повинен принести роботодавцю листок непрацездатності з вказанням причини непрацездатності «загальне захворювання – 1». Такий лікарняний оплачується за загальними правилами (50%, 60%, 70%, 100% середньої зарплати – залежно від страхового стажу). Тож, якщо працівник прямо не скаже і ніхто окремо не повідомить, то роботодавець не буде знати, на що саме хворів працівник. Якщо ж працівник знаходився на самоізоляції під медичним наглядом, то роботодавець отримає лікарняний лист, де у полі з причиною непрацездатності лікуючим лікарем буде зазначено «ізоляція від COVID-19 – 11». За таким листком непрацездатності лікарняні виплачують у розмірі 50% середньої зарплати (доходу) незалежно від страхового стажу. Виняток – медичні працівники. Їм допомога по тимчасовій непрацездатності у таких випадках виплачується в розмірі 100% середньої заробітної плати (доходу) незалежно від страхового стажу.

В разі хвороби (причина непрацездатності – 1), і в разі ізоляції (причина непрацездатності – 11) перші п'ять днів хвороби працівника оплачує за свій рахунок роботодавець, далі – ФСС. Також в обох випадках слід дотримуватися обмежень щодо суми лікарняних, якщо працівник за останні 12 місяців перед настанням страхового випадку заробив собі страховий стаж менше 6 місяців.

На підприємстві чи у ФОП працівник захворів на коронавірус. Звісно, перед роботодавцем постає питання щодо контактних осіб. Чи може (зобов'язаний) роботодавець наказати контактним особам піти на самоізоляцію? Ні, такого права (обов'язку) у нього наразі немає. Ба більше, не роботодавець визначає контактних осіб, а медичний заклад.

Все, що роботодавець може, це запропонувати працівникам, які контактували з хворим, попрацювати дистанційно (вдома) чи взяти відпустку, у тому числі «карантинну» відпустку без збереження зарплати. Якщо йдеться про щорічну відпустку – вносяться зміни до графіка відпусток. Якщо працівник не згоден, не допустити його до роботи лише через факт підозри, що він може бути носієм COVID-19, роботодавець не має права.

Чи має право роботодавець вимагати від працівника пройти тестування на COVID-19 та надати його результати перед виходом на роботу? Ні, не має такого права. Тестування є обов'язковим для:

- осіб, які звертаються за меддопомогою з ознаками COVID-19;
- осіб, які мали встановлений контакт з особою, хворою на COVID-19.

Також періодичне тестування проходять працівники медзакладів, які надають медичну допомогу хворим на COVID-19, працівники Нацгвардії та Нацполіції. Усі інші працівники не зобов'язані проходити тестування, навіть якщо їх роботодавцям цього дуже хочеться. Відмова від такого необов'язкового тестування не може бути підставою для притягнення до дисциплінарної відповідальності чи для відсторонення від роботи.

Працівник лояльний до роботодавця і пройшов тестування на COVID-19. Повідомляти чи ні про результат тестування – вирішує він сам. Інформація про стан здоров'я входить до персональних даних особи. Отже, потребує виконання правил щодо її обробки та зберігання. Звісно, частина роботодавців зберігає таку інформацію за працівниками певних категорій у зв'язку з необхідністю виконання приписів трудового законодавства. Проте ситуація з COVID-19 до цієї схеми не відноситься.

4.2 Підвищення стійкості роботи об'єктів господарської діяльності у воєнний час

Проведення заходів, спрямованих на підвищення стійкості роботи підприємств організацій і установ в умовах надзвичайних ситуацій – одно з основних завдань цивільного захисту.

Під стійкістю роботи промислового об'єкта, що безпосередньо виробляє якусь продукцію, розуміють його здатність в умовах надзвичайної ситуації випускати продукцію в запланованому обсязі та номенклатурі, а при одержанні слабких і середніх руйнувань або порушенні зв'язків по кооперації та поставкам відновлювати виробництво в мінімальний термін [127].

Під стійкістю роботи об'єктів, що безпосередньо не виробляють продукцію, розуміють їх здатність виконувати свої функції в умовах надзвичайної ситуації.

На стійкість роботи промислових об'єктів в умовах воєнної надзвичайної ситуації впливають наступні фактори:

- надійність захисту робітників та службовців від впливу зброї масового ураження;
- здатність інженерно-технічного комплексу об'єкта деякою мірою протистояти вражаючим факторам ядерного вибуху;
- захищеність об'єкта від вторинних вражаючих факторів;
- надійність системи постачання об'єкта всім необхідним для виробництва продукції (сировиною, паливом, комплектуючими виробами, електроенергією, водою, газом і т.п.);
- стійкість і безперервність управління виробництвом та цивільним захистом;
- підготовленість об'єкта до проведення рятувальних та інших невідкладних аварійно-відбудовних робіт і робіт по відбудові порушеного виробництва.

Перераховані фактори визначають основні, загальні для всіх об'єктів, шляхи підвищення стійкості роботи в особливий період (воєнний час) [128], а саме:

- забезпечення надійного захисту робітників та службовців від вражаючих факторів зброї масового ураження;
- захист основних виробничих фондів від вражаючих факторів, у тому числі і від вторинних;
- підвищення надійності та оперативності управління виробництвом;
- забезпечення стабільності постачання всім необхідним для випуску запланованої на воєнний час продукції;
- підготовка до відновлення порушеного виробництва.

Підвищення стійкості роботи об'єктів в умовах воєнної надзвичайної ситуації, досягається:

- завчасним проведенням комплексу інженерно-технічних, технологічних і організаційних заходів, спрямованих на максимальне зниження впливу вражаючих факторів зброї масового ураження;
- створенням умов для швидкої ліквідації наслідків надзвичайної ситуації.

Інженерно-технічні заходи включають комплекс робіт, що забезпечують підвищення стійкості виробничих будівель і споруд, верстатного і технологічного обладнання, комунально-енергетичних систем.

Технологічні заходи забезпечують підвищення стійкості роботи об'єкта шляхом зміни технологічного процесу, що сприяє прискоренню виробництва продукції та виключає можливість утворення вторинних вражаючих факторів. Організаційні заходи передбачають розробку та планування дій керівного складу, штабу, служб і формувань цивільного захисту при захисті робітників та службовців підприємств, проведенні рятувальних і невідкладних аварійно-відбудовних робіт, відновленні виробництва, а також випуску продукції на збережених потужностях.

ВИСНОВКИ

Коронавірус COVID-19, що швидко поширюється по всьому світу, був визнаний Всесвітньою організацією охорони здоров'я пандемією 11 березня 2020 р. В першому розділі кваліфікаційної роботи освітнього рівня «Магістр»:

- Описано стан предметної області аналітичного опрацювання відомостей щодо COVID-19.

- Проаналізовано видобування даних щодо COVID-19.

- Висвітлено попереднє опрацювання відомостей щодо COVID-19.

В другому розділі кваліфікаційної роботи:

- Досліджено методи видобування відомостей щодо COVID-19.

- Розглянуто збирання та попереднє опрацювання соціальних відомостей щодо COVID-19.

- Проаналізовано особливості інтеграції даних щодо COVID-19.

- Описано інструменти для видобування відомостей щодо COVID-19.

В третьому розділі кваліфікаційної роботи:

- Досліджено злиття та видобування відомостей щодо COVID-19 отриманих з ІюН-джерел.

- Описано використання штучного інтелекту для опрацювання відомостей щодо COVID-19.

- Проаналізовано методи навчання систем ШІ при діагностуванні COVID-19.

- Подано порівняльну характеристику методів навчання систем ШІ.

- Розглянуто методи глибинного навчання та COVID-19.

У розділі «Охорона праці та безпека в надзвичайних ситуаціях» проаналізовано нормативні документи, які визначають дії роботодавця, працівник якого захворів на COVID-19. Висвітлено питання підвищення стійкості роботи об'єктів господарської діяльності у воєнний час.

ПЕРЕЛІК ДЖЕРЕЛ

- 1 Abdulsalam Yassine, Shailendra Singh, Atif Alamri. “Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications”, *IEEE Access*, 2020.
- 2 Zhang, Qingguo, et al. "Multi-Source Medical Data Integration and Mining for Healthcare Services." *IEEE Access* 8 (2020): 165010-165017.
- 3 S. Din and A. Paul, “Smart health monitoring and management system: Toward autonomous wearable sensing for Internet of Things using big data analytics,” *Future Gener. Comput. Syst.*, vol. 111, p. 939, Feb. 2020.
- 4 N. C. Benda, T. C. Veinot, C. J. Sieck, and J. S. Ancker, “Broadband Internet access is a social determinant of health!,” *Amer. J. Public Health*, vol. 110, no. 8, pp. 1123–1125, Aug. 2020.
- 5 E. Silience, J. M. Blythe, P. Briggs, and M. Moss, “A revised model of trust in Internet-based health information and advice: Cross-sectional questionnaire study,” *J. Med. Internet Res.*, vol. 21, no. 11, Nov. 2019, Art. no. e11125.
- 6 K. Szulc and M. Duplaga, “The impact of Internet use on mental wellbeing and health behaviours among persons with disability,” *Eur. J. Public Health*, vol. 29, no. 4, pp. 185–425, Nov. 2019.
- 7 Centers for Disease Control and Prevention Mental Health and Coping During COVID-19|CDC. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html> (2020).
- 8 Kirzinger, A.; Kearney, A.; Hamel, L.; Brodie, M. *KFF Health Tracking Poll-Early April 2020: The Impact of Coronavirus on Life in America*; KFF: Oakland, CA, USA, 2020; pp. 1–30.
- 9 Zhou, C.; Su, F.; Pei, T.; Zhang, A.; Du, Y.; Lu, B.; Cao, Z.; Wang, J.; Yuan, W.; Zhu, Y.; et al. COVID-19: Challenges to GIS with Big Data. *Geogr. Sustain.* 2020, 1, 77–87. Available online: <https://www.sciencedirect.com/science/article/pii/S2666683920300092> (2020).

- 10 Mollalo, A.; Vahedi, B.; Rivera, K.M. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci. Total Environ.* 2020, 728, 138884.
- 11 Jahanbin, K.; Rahmanian, V. Using twitter and web news mining to predict COVID-19 outbreak. *Asian Pac. J. Trop. Med.* 2020, 26–28.
- 12 Coppersmith, G.; Dredze, M.; Harman, C. Quantifying Mental Health Signals in Twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Association for Computational Linguistics, Baltimore, MA, USA, 27 June 2014; pp. 51–60. Available online: <http://aclweb.org/anthology/W14-3207> (2020).
- 13 De Choudhury, M.; Counts, S.; Horvitz, E. Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013.
- 14 Feldman, R. Techniques and applications for sentiment analysis. *Commun. ACM* 2013, 56, 89.
- 15 Yu, Y.; Duan, W.; Cao, Q. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decis. Support Syst.* 2013, 55, 919–926.
- 16 Zhou, X.; Tao, X.; Yong, J.; Yang, Z. Sentiment analysis on tweets for social events. In Proceedings of the 17th International Conference on Computer Supported Cooperative Work in Design CSCWD, Whistler, BC, Canada, 27–29 June 2013; pp. 557–562.
- 17 Go, A.; Bhayani, R.; Huang, L. Twitter Sentiment Classification using Distant Supervision. *Processing* 2009, 1, 2009.
- 18 Pratama, B.Y.; Sarno, R. Personality classification based on Twitter text using Naïve Bayes, KNN and SVM. In Proceedings of the 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, Indonesia, 25–26 November 2015; pp. 170–174. Available online:

<https://www.semanticscholar.org/paper/Personality-classification-based-on-Twitter-text-PratamaSarno/6d8bf96e65b9425686bde3405b8601cc8c4f2779#references> (2020).

19 Chen, Y.; Yuan, J.; You, Q.; Luo, J. Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. In *Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018*; ACM Press: New York, NY, USA, 2018; pp. 117–125.

20 Barbosa, L.; Feng, J. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the Coling 2010 23rd International Conference on Computational Linguistics, Beijing, China, 23–27 August 2010*.

21 Pak, A.; Paroubek, P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *IJARCCCE* 2016, 5, 320–322.

22 Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Available online: <https://arxiv.org/abs/1603.04467> (2020).

23 Jin, C.; Zhang, S. Micro-blog Short Text Clustering Algorithm Based on Bootstrapping. In *Proceedings of the 12th International Symposium on Computational Intelligence and Design, Hangzhou, China, 14–15 December 2019*; pp. 264–266.

24 Bhosale, S.; Sheikh, I.; Dumpala, S.H.; Kopparapu, S.K. End-to-End Spoken Language Understanding: Bootstrapping in Low Resource Scenarios. *Proc. Interspeech* 2019, 1188–1192.

25 Mihalcea, R.; Banea, C.; Wiebe, J. Learning Multilingual Subjective Language via Cross-Lingual Projections. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 23–30 June 2007*; pp. 976–983.

- 26 Asghar, M.Z.; Ahmad, S.; Qasim, M.; Zahra, S.R.; Kundi, F.M. SentiHealth: Creating health-related sentiment lexicon using hybrid approach. *SpringerPlus* 2016, 5.
- 27 Duda, O; Kunanets, N; Matsiuk, O; Pasichnyk, V; Rzhеuskyi, A: Aggregation, Storing, Multidimensional Representation and Processing of COVID-19 Data. 2021 | book-chapter, DOI: 10.1007/978-3-030-63270-0_60.
- 28 Ogundele I.O, Popoola O.L, Oyesola O.O, Orija K.T, “A Review on Data Mining in Healthcare”, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* Volume 7, Issue 9, September 2018.
- 29 Duda O. et al., Mobile information system for monitoring the spread of viruses in smart cities. *Visnik Nacional'nogo universitetu "L'vivs'ka politehnika". Seriâ Informacijni sistemi ta mreži.* DOI: 10.23939/sisn2020.08.065.
- 30 Pooja H , Dr. Prabhudev Jagadeesh M P , “A Collective Study of Data Mining Techniques for the Big Health Data available from the Electronic Health Records”, *IEEE Xplore*, 2019.
- 31 Wencheng Sun, Zhiping Cai1, Fang Liu, Shengqun Fang, Guoyan Wang, “A Survey of Data Mining Technology on Electronic Medical Records”, *19th International Conference on eHealth Networking, Applications and Services*, 2017 IEEE.
- 32 Geetha Guttikonda, MadhaviLatha Pandala, Madhavi Katamaneni, “Diabetes Data Prediction Using Spark and Analysis in Hue Over Big Data”, *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)*.
- 33 Mario W. L. Moreira , Joel J. P. C. Rodrigues , Senior Member, Valery Korotaev, Jalal Al-Muhtadi, and Neeraj Kumar, “A Comprehensive Review on Smart Decision Support Systems for Health Care”, 2019 IEEE.
- 34 Sandeep Yadav, Aman Jain, Deepti Singh, “Early Prediction of Employee Attrition using Data Mining Techniques”, *IEEE*, 2018.

35 Alramzana Nujum Navaz, Elfadil Mohammed, Mohamed Adel Serhani and Nazar Zaki. "The Use of Data Mining Techniques to Predict Mortality and Length of Stay in an ICU", 12th International Conference on Innovations in Information Technology (IIT), 2016.

36 Pooja H., Dr. Prabhudev Jagadeesh M. P., A Collective Study of Data Mining Techniques for the Big Health Data available from the Electronic Health Records, 2019 IEEE.

37 D. Usha Rani, "A survey on Data Mining Tools and Techniques in Medical Field", International Journal of Advanced Networking & Applications (IJANA), Volume: 08, Issue: 05 Pages: 51-54 (2017) Special Issue, TECHSA-17.

38 Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* 2003, 3, 993–1022.

39 Ramage, D.; Manning, C.D.; Dumais, S. Partially labeled topic models for interpretable text mining. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011.

40 Passos, A.; Wallach, H.M.; Mccallum, A. Correlations and Anticorrelations in LDA Inference. In Proceedings of the Challenges in Learning Hierarchical Models: Transfer Learning and Optimization NIPS Workshop, Granada, Spain, 16–17 December 2011; pp. 1–5.

41 Yazdavar, A.H.; Al-Olimat, H.S.; Ebrahimi, M.; Bajaj, G.; Banerjee, T.; Thirunarayan, K.; Pathak, J.; Sheth, A. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, Australia, 31 July–3 August 2017; pp. 1191–1198.

42 Loper, E.; Bird, S. NLTK: The Natural Language Toolkit. arxiv.org. 2002. Available online: <http://portal.acm.org/citation.cfm?doid=1118108.1118117> (2020).

43 API Reference—TextBlob 0.16.0 Documentation. Available online: https://textblob.readthedocs.io/en/dev/api_reference.html#textblob.blob.TextBlob.sentiment (2020).

44 Ver Steeg, G.; Galstyan, A. Discovering structure in high-dimensional data through correlation explanation. *Adv. Neural Inf. Process Syst.* 2014, *1*, 577–585.

45 Li, Diya, Harshita Chaudhary, and Zhe Zhang. "Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining." *International Journal of Environmental Research and Public Health* 17.14 (2020): 4988.

46 Loper, E.; Bird, S. NLTK: The Natural Language Toolkit. In Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational linguistics, Philadelphia, PA, USA, 12 July 2002.

47 Ver Steeg, G.; Galstyan, A. Maximally informative hierarchical representations of high-dimensional data. *J. Mach. Learn. Res.* 2015, *38*, 1004–1012.

48 Ji, S.; Li, G.; Li, C.; Feng, J. Efficient interactive fuzzy keyword search. In Proceedings of the 18th International Conference World Wide Web. Association for Computing Machinery, New York, NY, USA, 20–24 April 2009; pp. 371–380.

49 A Fast, Offline Reverse Geocoder in Python. Available online: <https://github.com/thampiman/reversegeocoder> (2020).

50 Full List of US States and Cities. Available online: <https://github.com/grammakov/USA-cities-and-states> (2020).

51 Costa, Joao Pita, et al. "Meaningful Big Data Integration for a Global COVID-19 Strategy." *IEEE Computational Intelligence Magazine* 15.4 (2020): 51–61.

52 He, Yongqun, et al. "CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis." *Scientific Data* 7.1 (2020): 1-5.

53 Korn, Daniel, et al. "COVID-KOP: Integrating Emerging COVID-19 Data with the ROBOKOP Database." *Bioinformatics* (2020).

54 Guidotti, Emanuele, and David Ardia. "COVID-19 data hub." *Journal of Open Source Software* 5.51 (2020): 2376.

55 Ohannessian, Robin, Tu Anh Duong, and Anna Odone. "Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action." *JMIR public health and surveillance* 6.2 (2020): e18810.

56 Schwarz, Marius, et al. "COVID-19 and the academy: It is time for going digital." *Energy research & social science* 68 (2020): 101684.

57 Vidal-Alaball, Josep, et al. "Telemedicine in the face of the COVID-19 pandemic." *Atencion primaria* 52.6 (2020): 418.

58 Kandel, Nirmal, et al. "Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries." *The Lancet* (2020).

59 Fahey, Robert A., and Airo Hino. "COVID-19, digital privacy, and the social limits on data-focused public health responses." *International Journal of Information Management* 55 (2020): 102181.

60 Whitelaw, Sera, et al. "Applications of digital technology in COVID-19 pandemic planning and response." *The Lancet Digital Health* (2020).

61 T. Peng, Y. Lin, X. Yao, and W. Zhang, "An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data," *IEEE Access*, vol. 6, pp. 21924–21933, 2018.

62 H. Dai, Y. Ji, G. Yang, H. Huang, and X. Yi, "A privacy-preserving multi-keyword ranked search over encrypted data in hybrid clouds," *IEEE Access*, vol. 8, pp. 4895–4907, 2020.

63 T. V. Xuan Phuong, G. Yang, W. Susilo, F. Guo, and Q. Huang, “Sequence aware functional encryption and its application in searchable encryption,” *J. Inf. Secur. Appl.*, vol. 35, pp. 106–118, Aug. 2017.

64 Z. Xia, X. Wang, X. Sun, and Q. Wang, “A secure and dynamic multikeyword ranked search scheme over encrypted cloud data,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 340–352, Feb. 2016.

65 M. He, M. Chang, and X. Wu, “A collaborative filtering recommendation method based on differential privacy,” *J. Comput. Res. Develop.*, vol. 54, no. 7, pp. 1439–1451, 2017.

66 T. Wang and S. He, “An improved collaborative filtering recommendation algorithm with differentially privacy,” *Inf. Secur. Technol.*, vol. 7, no. 4, pp. 26–28, 2016.

67 Z. Xian, Q. Li, X. Huang, J. Lu, and L. Li, “Differential privacy protection for collaborative filtering algorithms with explicit and implicit trust,” *Acta Electronica Sinica*, vol. 46, no. 12, pp. 3050–3059, 2018.

68 C. Yin, L. Shi, R. Sun, and J. Wang, “Improved collaborative filtering recommendation algorithm based on differential privacy protection,” *J. Supercomput.*, vol. 76, no. 7, pp. 5161–5174, Jul. 2020.

69 Y. Xiao, L. Xiong, S. Zhang, and Y. Cao, “LocLok: Location cloaking with differential privacy via hidden Markov model,” *Proc. VLDB Endowment*, vol. 10, no. 12, pp. 1901–1904, Aug. 2017.

70 S. Yang, J. Xu, X. Yang, and X. Ren, “Bayesian network-based highdimensional crowdsourced data publication with local differential privacy,” *Scientia Sinica Informationis*, vol. 49, no. 12, pp. 1586–1605, Dec. 2019.

71 J. Wang, Z. Cai, Y. Li, D. Yang, J. Li, and H. Gao, “Protecting query privacy with differentially private k-anonymity in location-based services,” *Pers. Ubiquitous Comput.*, vol. 22, no. 3, pp. 453–469, Jun. 2018.

72 S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, “A caching and spatialK-anonymity driven privacy enhancement scheme in continuous location-based services,” *Future Gener. Comput. Syst.*, vol. 94, pp. 40–50, May 2019.

73 F. Casino, J. Domingo-Ferrer, C. Patsakis, D. Puig, and A. Solanas, “A k-anonymous approach to privacy preserving collaborative filtering,” *J. Comput. Syst. Sci.*, vol. 81, no. 6, pp. 1000–1011, Sep. 2015.

74 P. Zhao, J. Li, F. Zeng, F. Xiao, C. Wang, and H. Jiang, “ILLIA: Enabling k -anonymity-based privacy preserving against location injection attacks in continuous LBS queries,” *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1033–1042, Apr. 2018.

75 S.-H. Wang, Y. Zhang, Y.-J. Li, W.-J. Jia, F.-Y. Liu, M.-M. Yang, and Y.-D. Zhang, “Single slice based detection for Alzheimer’s disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization,” *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10393–10417, May 2018.

76 S. Wang, J. Sun, I. Mehmood, C. Pan, Y. Chen, and Y. Zhang, “Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling,” *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, Jan. 2020, Art. no. e5130.

77 Y.-D. Zhang, V. V. Govindaraj, C. Tang, W. Zhu, and J. Sun, “High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model,” *J. Med. Imag. Health Informat.*, vol. 9, no. 9, pp. 2012–2021, Dec. 2019.

78 Y. Zhang, S. Wang, Y. Sui, M. Yang, B. Liu, H. Cheng, J. Sun, W. Jia, P. Phillips, and J. M. Gorriz, “Multivariate approach for Alzheimer’s disease detection using stationary wavelet entropy and predator-prey particle swarm optimization,” *J. Alzheimer’s Disease*, vol. 65, no. 3, pp. 855–869, Sep. 2018.

79 C. Kang, X. Yu, S.-H. Wang, D. Guttery, H. Pandey, Y. Tian, and Y. Zhang, “A heuristic neural network structure relying on fuzzy logic for images

scoring,” *IEEE Trans. Fuzzy Syst.*, early access, Jan. 13, 2020, doi: 10.1109/TFUZZ.2020.2966163.

80 S.-H. Wang, Y.-D. Zhang, M. Yang, B. Liu, J. Ramirez, and J. M. Gorriz, “Unilateral sensorineural hearing loss identification based on doubledensity dual-tree complex wavelet transform and multinomial logistic regression,” *Integr. Comput.-Aided Eng.*, vol. 26, no. 4, pp. 411–426, Sep. 2019.

81 Bindiya M K, Dr. Ravikumar, Dr. Mohan H. S., “Monitoring the appliances used in health care using Medical Big Data”, Proceedings of the International Conference on Inventive Computation Technologies(ICICT-2018).

82 M. Durairaj, V. Ranjani, “Data Mining Applications In Healthcare Sector: A Study”, *International Journal of Scientific & Technology Research*” Volume 2, October 2013.

83 WEKA. The workbench for machine learning, <https://www.cs.waikato.ac.nz/ml/weka/>.

84 KEEL. ASSET DATA MANAGEMENT & ASSET INTEGRITY SERVICES, <https://keelsolution.com/>.

85 R for Data Science: Welcome, <https://r4ds.had.co.nz/>.

86 KNIME | Open for Innovation, <https://www.knime.com/>.

87 RapidMiner | Best Data Science & Machine Learning Platform, <https://rapidminer.com/>.

88 Orange Data Mining - Data Mining, <https://orange.biolab.si/>.

89 WS-DREAM. Web Service QoS Datasets, <http://inpluslab.com/wsdream/>.

90 Murdoch TB and Detsky AS. The inevitable application of big data to health care. *JAMA* 309: 1351–2, 2013.

91 Kolker E, Özdemir V, and Kolker E. How Healthcare can refocus on its Super-Customers (Patients, $n = 1$) and Customers (Doctors and Nurses) by Leveraging Lessons from Amazon, Uber, and Watson. *OMICS* 20:329–33, 2016.

92 Dilsizian SE and Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr Cardiol Rep*, 16:441, 2014.

93 Xu X, Chen P, and Wang J. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. DOI: 10.1007/s11427-0201637-5, 2020.

94 Huang C, Wang Y and Li X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. Doi: 10.1016/S01406736(20)30183-5, 2020.

95 Chan JF-W, Yuan S and Kok K-H. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. Doi: 101016/S01406736(20)30154-9, 2020.

96 Latif J, Xiao C, Imran A. and Tu S. Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review, 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, pp. 1-5, 2019.

97 Yang Y. and Chen T. Analysis and Visualization Implementation of Medical Big Data Resource Sharing Mechanism Based on Deep Learning, in *IEEE Access*, vol. 7, pp. 156077-156088, 2019.

98 Zhao H, Li G. and Feng W. Research on Application of Artificial Intelligence in Medical Education, 2018 International Conference on Engineering Simulation and Intelligent Control (ESAIC), Changsha, 2018, pp. 340-342.

99 Al-Turjman F, Nawaz M.H. and Ulusar U.D. Intelligence in the Internet of Medical Things era: A systematic review of current and future trends, *Computer Communications*, doi: <https://doi.org/10.1016/j.comcom.2019.12.030>, (2019).

100 Chan, Yung-Kuan, Yung-Fu Chen, Tuan Pham, Weide Chang, and Ming-Yuan Hsieh. 2018. Artificial Intelligence in Medical Applications. *Journal of Healthcare Engineering*, 4827875. <https://doi.org/10.1155/2018/4827875>.

101 Administration UFaD. Guidance for industry: electronic source data in clinical investigations. [Online] Available: <https://www.fda.gov/downloads/drugs/guidances/ucm328691>.

102 Hussain, Adedoyin Ahmed, et al. "AI techniques for COVID-19." *IEEE Access* 8 (2020): 128776-128795.

103 Gillies RJ, Kinahan PE, Hricak and H. Radiomics. Images are more than pictures, they are data. *Radiology*, 278:563–77, 2016.

104 Li CY, Liang GY and Yao WZ. Integrated analysis of long noncoding RNA competing interactions reveals the potential role in progression of human gastric Cancer. *Int J Oncol* 48:1965–76, 2016.

105 Shin H, Kim KH and Song C. Electro-diagnosis support system for localizing neural injury in an upper limb. *J Am Med Inform Assoc*, 17:345–7, 2010.

106 Karakulah G, Dicle O and Koşaner O. Computer based extraction of phenotypic features of human congenital anomalies from the digital literature with natural language processing techniques. *Stud Health Technology Inform*, 205:570–4, 2014.

107 Becky McCall. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Published by Elsevier Ltd. Doi: 10.1016/S2589-7500(20)30054-6. 2020.

108 Commissioner O of the. Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices. FDA. [Online] Available: <https://news-events/press-announcements/statement-fdacommissioner-scott-gottlieb-mdsteps>, 2019.

109 Al-Turjman F. Intelligence and security in big 5G-oriented IoNT: An overview. Elsevier Future Generation Computer Systems 102, 357–368 2019.

110 Bedford J, Enria D, Giesecke J, Heymann D.L, Ihekweazu C, Kobinger V, Lane H. C, Memish V, Oh M.D and Schuchat A. COVID-19: towards controlling of a pandemic, Lancet 395 (10229) (2020) 1015– 1018, 2020.

111 Chollet F. Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258, 2017.

112 Matar G, Lina J.-M, and Kaddoum G. Artificial neural network for inbed posture classification using bed-sheet pressure sensors, IEEE Journal of biomedical and health informatics, no. 10, pp. 101–110, 2020.

113 Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, Bernheim A and Siegel E. Rapid AI development cycle for the Coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv preprint arXiv:2003.05037. 2020.

114 Anurag Ranjan, Varun Jampani, Lukas Balles, Kihwan Kim, Deqing Sun, Jonas Wulff, and Michael J Black. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 12240–12249, 2019.

115 Anurag Ranjan, Varun Jampani, Lukas Balles, Kihwan Kim, Deqing Sun, Jonas Wulff, and Michael J Black. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 12240–12249, 2019.

116 Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and

egomotion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 989–997, 2019.

117 Al-Turjman F., Mohammed Zaki Hasan and Hussain Al-Rizzo. Task scheduling in cloud-based survivability applications using swarm optimization in IoT. Wiley special issue article, 2018.

118 Wan KH, Huang SS, Young A and Lam DS. Precautionary measures needed for ophthalmologists during pandemic of the coronavirus disease 2019 (COVID-19). *Acta Ophthalmol*, 2020.

119 Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q and Cao K. Artificial intelligence distinguishes COVID-19 from community-acquired pneumonia on chest CT. *Radiology*: 200905, 2020.

120 E. M. Forman, S. G. Kerrigan, M. L. Butryn, A. S. Juarascio, S. M. Manasse, S. Ontañón, D. H. Dallal, R. J. Crochiere, and D. Moskow. Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss?. *Journal of Behavioral Medicine*, pp. 1–15, 2018.

121 Wim Naude. Artificial Intelligence against COVID-19: An Early Review. [Online] Available: <https://towardsdatascience.com/artificialintelligence-against-covid-19-an-early-review-92a8360edaba>. 2020.

122 Max Smolaks. Deep learning systems can speed up diagnosis of Covid-19 patients. [Online] Available: <https://aibusiness.com/deep-learning-systems-can-identify-covid-19-infections-on-x-rays/>. 2020.

123 Закон України Про захист населення від інфекційних хвороб (Відомості Верховної Ради України (ВВР), 2000, № 29, ст. 228), <https://zakon.rada.gov.ua/laws/show/1645-14#Text>.

124 Стаття 44-3. Порушення правил щодо карантину людей, https://protocol.ua/ua/kodeks_ukraini_pro_administrativni_pravoporushennya_stat_tya_44_3_1/.

125 Кабінет міністрів України. Постанова від 17 квітня 2019 р. № 337. Про затвердження Порядку розслідування та обліку нещасних випадків, професійних захворювань та аварій на виробництві, <https://zakon.rada.gov.ua/laws/show/337-2019-%D0%BF#Text>.

126 Державна служба України з питань праці, <https://dsp.gov.ua/>.

127 Конспект лекцій дисципліни «Цивільний захист і охорона праці в галузі», змістовний модуль «Цивільний захист», для студентів усіх спеціальностей та всіх форм навчання / Укл.: М. О. Журавель – Запоріжжя: НУ «Запорізька політехніка». Каф. ОП і НС, 2020 р. – 49 с.

128 Стручок, Володимир Сергійович, Олена Степанівна Стручок, and Дарія Володимирівна Мудра. "Навчальний посібник до написання розділу дипломного проекту та дипломної роботи "Безпека в надзвичайних ситуаціях "для студентів всіх спец. денної, заочної (дистанційної) та екстернатної форм навчання." (2017).

ДОДАТКИ

Тези конференції

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ**

МАТЕРІАЛИ

VIII НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

**«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»**



9–10 грудня 2020 року

**ТЕРНОПІЛЬ
2020**

Б. Гнатків, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА ДЛЯ ЗАБЕЗПЕЧЕННЯ ФУНКЦІОНАЛЬНИХ МОЖЛИВОСТЕЙ КОНЦЕПЦІЇ «РОЗУМНЕ МІСТО»	
V. Hnatkiv, N. Kunanets INFORMATION SYSTEM FOR PROVIDING FUNCTIONAL POSSIBILITIES OF THE "SMART CITY" CONCEPT	77
Д. Манько, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА ЕЛЕКТРОННОГО ГОЛОСУВАННЯ З КРИПТОГРАФІЧНИМ ЗАХИСТОМ ДАНИХ	
D. Manko, N. Kunanets ELECTRONIC VOTING INFORMATION SYSTEM WITH CRYPTOGRAPHIC DATA PROTECTION	78
П. Місюрка, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА ОРГАНІЗАЦІЇ ДОЗВІЛЛЯ З ВРАХУВАННЯМ ЕТНІЧНИХ ОСОБЛИВОСТЕЙ РЕГІОНУ	
P. Misyurka, N. Kunanets INFORMATION SYSTEM OF LEISURE ORGANIZATION TAKING INTO ACCOUNT THE ETHNIC PECULIARITIES OF THE REGION	79
С. Сем'янчук, Т. Шестакевич, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА СУПРОВОДУ СОЦІАЛЬНИХ ПРОЕКТІВ	
S. Semyanchuk, T. Shestakevich, N. Kunanets INFORMATION SYSTEM OF SOCIAL PROJECT SUPPORT	80
А. Юськевич, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА РОЗВИТКУ ТЕРИТОРІАЛЬНИХ ГРОМАД	
A. Yuskevich, N. Kunanets TERRITORIAL COMMUNITY DEVELOPMENT INFORMATION SYSTEM	81
А. Ванік, В. Гніздюх, Т. Масєвський АНАЛІТИЧНЕ ОПРАЦЮВАННЯ ВІДОМОСТЕЙ ЩОДО COVID-19	
A. Vanyuk, V. Hnizdiukh, O. Yaiechnyk, T. Maievskiy ANALYTICAL PROCESSING OF COVID-19 INFORMATION	82
А. Ванік, О. Пригоцький, О. Яєчник., Т. Масєвський ВИКОРИСТАННЯ ІОТ-ПРИСТРОЇВ ДЛЯ ВІДБОРУ БІОМЕДИЧНИХ ДАНИХ В УМОВАХ ПАНДЕМІЇ COVID-19	
A. Vanyuk, O. Prytotskiy, O. Yaiechnyk, T. Maievskiy USE OF IOT DEVICES FOR BIOMEDICAL DATA SELECTION IN A COVID- 19 PANDEMIC	83
І. Дурибаба, Н. Кунаець ІНФОРМАЦІЙНА СИСТЕМА ДИСТАНЦІЙНОГО КОНСУЛЬТУВАННЯ ТА ОНЛАЙН ЗАПИСУ ДЕРМАТОЛОГІЧНОГО ЦЕНТРУ	
I. Durybaba, N. Kunanets THE INFORMATION SYSTEM FOR REMOTE CONSULTATION AND ONLINE RECORDING OF THE DERMATOLOGICAL CENTRE	84
Я. Ватаг, А. Василюк, Н. Кунаець СТВОРЕННЯ СИСТЕМИ НАДАННЯ РЕКОМЕНДАЦІЙ З ВИБОРУ РОЗВАЖАЛЬНИХ ЗАКЛАДІВ МІСТА ЛЬВОВА	
J. Vatag, A. Vasyliuk, N. Kunanets CREATION OF A SYSTEM OF PROVIDING RECOMMENDATIONS FOR THE CHOICE OF ENTERTAINMENT FACILITIES OF THE CITY OF LVIV	85
А. Крашівський РОЗРОБКА ВЕБ-СИСТЕМИ З ВИКОРИСТАННЯМ NODE.JS ТА MONGODB НА ПРИКЛАДІ СИСТЕМИ АВТОМАТИЗАЦІЇ HR-ПРОЦЕСІВ	
A. Krashivskiy WEB SYSTEM DEVELOPMENT USING NODE.JS AND MONGODB ON EXAMPLES OF HR-PROCESS AUTOMATION SYSTEM	86

УДК 004.62

Ваник А.Г., Гніздюх В.Г., Яєчник О.П., Маєвський Т.О.

(Тернопільський національний технічний університет імені Івана Пулюя)

АНАЛІТИЧНЕ ОПРАЦЮВАННЯ ВІДОМОСТЕЙ ЩОДО COVID-19

UDC 004.62

Vanyk A.H., Hnizdiukh V.H., Yaiechnyk O.P., Maievskiy T.O.

ANALYTICAL PROCESSING OF COVID-19 INFORMATION

Глобальна пандемія Covid-19 вимагає комплексної та глобальної реакції всіх світових та національних медичних організацій та установ, що функціонують в галузі охорони здоров'я. Covid-19 спричинив загострення проблем в галузі охорони здоров'я та виявив необхідність безперешкодного, швидкого та своєчасного обміну даними щодо глобальних пандемій та підвищив вимоги щодо оперативного реагування [1]. Оскільки COVID-19 швидко поширився по всьому світу, ефективне використання моделей прогнозування може відіграти визначну роль для допомоги в управлінні ресурсами охорони здоров'я та плануванні профілактичних заходів.

Алгоритми та методи аналітичного опрацювання даних – це добре відомі інструменти та засоби для розроблення прогнозних моделей та практичного аналізу даних. З їх використанням можна видобувати приховану або неявно подану корисну інформацію з наборів та колекцій необроблених даних [2]. Видобуті знання та відомості щодо глобальної пандемії COVID-19 можуть бути використані не тільки в галузі охорони здоров'я а й у різних сферах. На даний час в галузі охорони здоров'я створено та продовжується накопичення великої кількості даних щодо COVID-19, включаючи дані про пацієнтів, супутні захворювання та діагнози.

В інтелектуальному аналізі відомостей щодо COVID-19, виділяють дві категорії завдань. Перша категорія – це описові завдання, що стосуються загальних властивостей даних про COVID-19. Друга категорія – це передбачувальні (прогнозні) завдання, основною метою яких є побудова моделей, що можуть оцінити відображення корисних знань від інформаційних входів до виходів за допомогою навчальної вибірки даних. Навчені моделі можуть бути використані для прогнозування результатів для наборів вхідних відомостей щодо COVID-19. У порівнянні з традиційним статистичним аналізом, методи, що відносяться до другої категорії будуть більш гнучкими та ефективними в задачах дослідницького аналізу [3].

Це лише початок наукових досліджень щодо аналітичного опрацювання відомостей, зібраних про COVID-19 з різнотипових джерел. Незважаючи на те, що сформовані на даний час прогнозні моделі не дуже точні [2], вони можуть бути корисними для побудови точних моделей на основі більшої агрегації даних щодо COVID-19. Відставання у прогнозуванні може бути наслідком неоднозначності захворюваності в різних країнах. Подальші дослідження потребують поглибленого аналізу доступних наукових джерел.

Література.

1. Radanliev, Petar, David De Roure, and Rob Walton. "Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development-In the first wave of the Covid-19 pandemic." *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14.5 (2020): 1121-1132.
2. Ayyoubzadeh, Seyed Mohammad, et al. "Predicting COVID-19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study." *JMIR Public Health and Surveillance* 6.2 (2020): e18828.
3. Sherstinsky, Alex. "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network." *Physica D: Nonlinear Phenomena* 404 (2020): 132306.

УДК 004.67

Ваник А.Г., Притоцький О.О., Яєчник О.П., Маєвський Т.О.
(Тернопільський національний технічний університет імені Івана Пулюя)

ВИКОРИСТАННЯ ІОТ-ПРИСТРОЇВ ДЛЯ ВІДБОРУ БІОМЕДИЧНИХ ДАНИХ В УМОВАХ ПАНДЕМІЇ COVID-19

UDC 004.67

Vanyk A.H., Prytotskyi O.O., Yaiechnyk O.P., Maievskiy T.O.

USE OF IOT DEVICES FOR BIOMEDICAL DATA SELECTION IN A COVID- 19 PANDEMIC

Повідомляється, що в світі до завершення 2020 року буде підключено понад 50 мільярдів пристроїв з використанням засобів радіозв'язку [1]. На їх основі формуються IoT-мережі давачів, мобільних пристроїв, радіоідентифікаційних міток та виконавчих пристроїв, котрі запрограмовані на збирання даних із середовища користувача. IoT-мережі ефективно використовуються у багатьох галузях, зокрема в системі охорони здоров'я, управлінні процесами постачання енергоносіїв та комунальних послуг, розумних будинках, безпекових системах та сільському господарстві. Функціональні можливості IoT-пристроїв при комплексному використанні з «розумними» інформаційними технологіями суттєво розширюють можливості надання високоякісних та своєчасних послуг в умовах глобальної пандемії COVID-19. Послуги сформовані на базі IoT-пристроїв із залученням смартфонів стали інноваційною мережевою парадигмою яка консолідує розподілені послуги та фізичні об'єкти.

В роботі [2] подано опис системи виявлення та моніторингу COVID-19 у режимі реального часу. Запропонована авторами система використовує інформаційно-технологічний концепт Інтернету речей (IoT) для відбору відомостей щодо симптомів COVID-19, раннього виявлення підозр захворювання, моніторингу реакції на лікування інфікованих громадян, постлікувального спостереження пацієнтів та розширення розуміння природи захворювання. В [3] Ндіає описує вплив глобальної пандемії COVID-19, на розвиток інформаційних та комунікаційних технологій, зокрема IoT. Він розглядає внесок IoT та пов'язаних з ними сенсорних технологій у процеси відстеження вірусів та пом'якшення наслідків. В публікації розглядаються супутні проблеми розгортання апаратного забезпечення давачів в умовах швидко поширюваної пандемії. Сінх [4] досліджує загальне застосування IoT, пропонуючи перспективну дорожню карту для подолання пандемії COVID-19. Автор аналізує дванадцять програм для IoT-пристроїв.

Очікується, що світові потрібно буде боротися з пандемією COVID-19 з використанням обережних заходів, поки не буде розроблена дієва вакцина. Тому формування ефективних інформаційно-технологічних систем для відбору біомедичних даних з використанням IoT-пристроїв в умовах пандемії COVID-19 є актуальним напрямком досліджень та потребує детальнішого опрацювання.

Література.

1. Kolhar, Manjur, et al. "A three layered decentralized IoT biometric architecture for city lockdown during COVID-19 outbreak." *IEEE Access* 8 (2020): 163608-163617.
2. Otoom, Mwaffaq, et al. "An IoT-based framework for early identification and monitoring of COVID-19 cases." *Biomedical Signal Processing and Control* 62 (2020): 102149.
3. Ndiaye, Musa, et al. "IoT in the Wake of COVID-19: A Survey on Contributions, Challenges and Evolution." *IEEE Access* 8 (2020): 186821-186839.
4. Singh, Ravi Pratap, et al. "Internet of things (IoT) applications to fight against COVID-19 pandemic." *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* (2020).