

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

МЕЛЬНИКОВА Наталія Іванівна

УДК 004.89: 002.53

**МОДЕЛІ ТА МЕТОДИ ПІДТРИМКИ ПЕРСОНАЛІЗОВАНИХ РІШЕНЬ У
МЕДИЧНИХ СИСТЕМАХ**

05.13.23. - системи та засоби штучного інтелекту

РЕФЕРАТ

дисертації на здобуття наукового ступеня
доктора технічних наук

Львів-2023

Дисертацією є рукопис.

Робота виконана в Національному університеті «Львівська політехніка».

Науковий консультант

доктор технічних наук, професор

ШАХОВСЬКА Наталя Богданівна,

Національний університет «Львівська політехніка»,

завідувач кафедри «Системи штучного інтелекту».

Офіційні опоненти:

доктор технічних наук, професор, академік АН ВШ України, **Зайченко Юрій Петрович**, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», професор кафедри математичних методів системного аналізу;

доктор технічних наук, професор, **Березький Олег Миколайович**, Західноукраїнський національний університет, професор кафедри комп'ютерної інженерії;

доктор технічних наук, професор, **Субботін Сергій Олександрович**, Запорізький національний технічний університет, завідувач кафедри програмних засобів.

Захист відбудеться *« 3 » березня 2023 року о 14⁰⁰ годині* на засіданні спеціалізованої вченої ради Д 35.052.14 у Національному університеті «Львівська політехніка» за адресою: 79013, м.Львів, вул. С. Бандери, 12, аудиторія 226 головного корпусу.

З дисертацією можна ознайомитись у науково-технічній бібліотеці Національного університету «Львівська політехніка» за адресою: 79013, м.Львів, вул. Професорська, 1.

Реферат розісланий *« 09 » лютого 2023 року*.

*Вчений секретар
спеціалізованої вченої ради
Д 35.052.14*

А. Є. Батюк

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Персоналізована медицина з даними орієнтованими на пацієнта набуває все більшого значення з багатьох причин, серед яких: підвищення точності процесу обробки медичних даних для пошуку персоналізованих рішень, що впливає на забезпечення якісного та ефективного медичного обслуговування; можливість прогнозування зміни стану конкретного пацієнта тощо.

За результатами дослідження персональних даних пацієнта з точки зору протоколу лікування характеризуються викидами, шумами, пропусками даних, часовою залежністю та мультимодальністю. Важливим фактором є потреба опрацювати інформацію пацієнта, опираючись на протоколи хвороб, не лише однієї хвороби з якою звернувся хворий, але й усіх супутніх захворювань та індивідуальних особливостей його стану.

Існуючі прикладні системи, орієнтовані на забезпечення персоналізації рішень, в рамках виду робіт при опрацюванні індивідуальних медичних даних пацієнтів. Особливості таких систем характеризуються процесом прогнозування, яке спрямоване лише для одного пацієнта, а не для групи пацієнтів з однаковими параметрами, які у часі можуть змінюватися під впливом різних факторів, що значно знижує якість прийнятих персоналізованих медичних рішень.

На сьогодні існуючі підходи та рішення характеризуються стійкістю до шумів, розпаралельністю обробки даних, високою точністю і стабільністю, визначенням важливих ознак, знаходженням нелінійно розділених кластерів, високою чутливістю та специфічністю у прогнозуванні, але це стосується лише класичних методів машинного навчання, які використовуються для вирішення такого типу задач, та які дають високі показники лише при розв'язанні окремих задач обробки персональних даних. Для комплексного підходу необхідно введення евристик, припущення про незалежність параметрів, робота не лише з бінарними ознаками об'єктів, пошук асоціативних залежностей з малою підтримкою, пошук оптимального рішення, підвищення точності прийняття лікарських рішень.

Наступна характерна особливість медичних даних – це або малі набори даних (сотні екземплярів даних), наприклад, інформація про пацієнтів з орфанними хворобами, або великі набори даних (тисячі та сотні тисяч екземплярів даних), наприклад для пандемій, таких, як COVID - 19. Іноді, залежно від типу захворювання та особливостей пацієнта кількість параметрів також може змінюватися від одиниць до сотень. При роботі з медичними даними, виникають проблеми як для малих наборів даних, так і для великих, а саме:

- для малих – низька повторюваність результатів під час пошуку персоналізованих рішень;
- для великих – необхідність відбору важливих ознак та висока обчислювальна складність.

Отже, опрацювання універсальними методами великих та малих наборів мультимодальних даних не дає змоги вирішити низку проблем, а саме:

- залежність якості моделі машинного навчання від результатів попередньої обробки даних для великих наборів даних;
- підвищення повторюваності результатів під час пошуку персоналізованих рішень на малих наборах даних;
- забезпечення генералізації при опрацюванні малої вибірки медичних даних

слабкими предикторами.

Усе перераховане вище зумовлює потребу у розробленні нових методів пошуку персоналізованих рішень, які повинні враховувати наявність факту мультимодальності даних, необхідність заповнення даних, залежності між параметрами, забезпечувати точність прогнозування цільових даних. Існуючі методи штучного інтелекту вирішують лише окремі підзадачі опрацювання медичних даних.

Отже, проблема розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізація результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта є актуальною науково-прикладною проблемою.

Зв'язок роботи з науковими програмами, планами та темами. Дисертаційна робота відповідає науковому напрямку кафедри системи штучного інтелекту Національного університету, а саме: «Методи та засоби обробки, консолідації та аналізу персоналізованої медичної інформації», та є складовою частиною проєктів, які виконувалися в межах дербюджетних науково-дослідних робіт та грантових робіт: ДБ «Інформаційна технологія опрацювання персоналізованої медичної інформації» № 0119U002257 (керівник ДБ); ДБ «Технології та системи оброблення і зберігання персоналізованих військових медичних даних» № 0121U107809; ДБ «Розроблення інформаційної технології оцінювання та прогнозування надійності програмного забезпечення методами машинного навчання» № 0121U109527; Госпдоговір ТОВ «БІОФАРМА ПЛАЗМА», спрямований на розроблення інформаційної системи обліку та аналізу пацієнтів з орфанними хворобами «Реєстр первинних імунодефіцитів»; Проєкт Національного Фонду Досліджень України «Система підтримки прийняття рішень моделювання поширення вірусних інфекцій» № 211/01.220; Грант Central European Initiatives Extraordinary call «STOP COVID - 19»; Грант CELTIC EUROGIA PROPOSAL «Integrated care for next generation iCare4NextG».

Мета і завдання дослідження. Метою дисертаційної роботи є розроблення моделей, методів машинного навчання та засобів для підвищення прогнозованої точності та візуалізації результатів опрацювання персональних даних щодо оцінки стану хворого, що забезпечить якісний процес прийняття персоналізованих медичних рішень.

Мета дисертаційної роботи визначає необхідність вирішення таких задач для опрацювання персоналізованих медичних даних:

1. Провести аналіз проблеми опрацювання мультимодальних персоналізованих медичних даних і попередньої обробки даних та аналізу якості моделі даних;
2. Розробити інформаційну модель стану пацієнта у багатовимірному просторі умов, яка за рахунок додаткових вимірів та параметрів забезпечить прогнозування цільових змінних;
3. Розробити метод двоетапної обробки малих наборів даних, що забезпечуватиме підвищення точності процесу узагальнення результатів на малих наборах даних;
4. Розробити метод групування моделей машинного навчання, який забезпечить прогнозування даних та паралелізацію процесу обробки даних при умові консолідації результатів;
5. Розробити метод зменшення розмірності вхідних даних, що забезпечить підвищення точності вибору пріоритетних ознак на великих наборах даних;
6. Розробити метод заповнення пропусків даних для підвищення стійкості моделі до

- помилки даних та паралелізації обчислень;
7. Розробити метод пошуку зміни стану пацієнта з урахуванням простору умов та аналізу характеристик його стану, що забезпечує підвищення точності підбору схеми лікування;
 8. Розробити метод консолідації мультимодальних даних для забезпечення агрегації даних з різною структурою;
 9. Розробити і апробувати програмні модулі щодо підтримки прийняття персоналізованих рішень на основі даних.

Об'єктом дослідження є процес обробки та аналізу персоналізованих медичних даних.

Предметом дослідження є методи прийняття рішень, методи штучного інтелекту, зокрема машинного навчання для аналізу персоналізованих медичних даних.

Методи дослідження. Для досягнення поставленої мети використано: методи прийняття рішень – для забезпечення чіткості та направленості та у пошуку рішень стосовно вибору цільових схем лікування; теорію реляційних баз даних – для відображення простору умов щодо ідентифікації стану хворого та урахування додаткових вимірів та параметрів; систему операцій реляційної алгебри – для формалізації залежностей між множинами параметрів пацієнта та показників оцінки його стану; методи штучного інтелекту – для виявлення закономірностей та залежностей між цільовими змінними, кластеризації груп пацієнтів за відповідними ознаками, для усунення невизначеностей у просторі даних, для прогнозування цільових змінних; методи об'єктно-орієнтованого аналізу і проектування – для визначення семантичних зв'язків між джерелами даних; методи консолідації даних – для агрегації даних з різною структурою; методи моделювання інформаційних систем.

Наукова новизна роботи. У дисертаційній роботі вирішено важливу науково-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізації результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

Отримано такі нові наукові результати:
вперше

- розроблено модель відображення стану пацієнта в багатовимірному просторі умов, що за рахунок додаткових вимірів та параметрів забезпечила підвищення точності прогнозування;
- розроблено метод двоетапної обробки малих наборів даних на основі ієрархічного предиктора, який за рахунок кластеризації та прогнозування забезпечує підвищення точності опрацювання нових вхідних наборів даних;
- розроблено метод групування моделей машинного навчання, який, враховуючи основні ознаки та повторне навчання, забезпечує підвищення точності прогнозування даних;
- розроблено метод зменшення розмірності вхідних даних, який завдяки класифікаторам, асоціативним правилам та узагальненому рангу ознак на основі індексу Жакара, забезпечує підвищення точності вибору пріоритетних ознак на великих наборах даних;

удосконалено:

- метод заповнення пропусків даних, який, використовуючи асоціативні та

- продукційні правила, забезпечує підвищення стійкості моделі до помилок даних;
- метод пошуку зміни стану пацієнта, який, використовуючи простір умов та аналіз характеристик стану, забезпечує підвищення точності підбору схеми лікування;
- метод консолідації мультимодальних даних, який за рахунок попереднього визначення структури та семантики даних забезпечує агрегування даних з різною структурою;

набула подальшого розвитку:

- теорія підтримки прийняття персоналізованих рішень, яка ґрунтується на розробленні моделі відображення стану пацієнта в багатовимірному просторі умов та розроблених методах двоетапної обробки малих наборів даних, групуванні моделей машинного навчання, зменшенні розмірності вхідних даних, заповненні пропусків даних, пошуку зміни стану пацієнта та консолідації мультимодальних даних, що забезпечило адаптацію прийняття персоналізованих медичних рішень.

Практичне значення одержаних результатів та розроблених методів полягає у тому, що вони є структурною складовою процесу підтримки прийняття рішень з урахуванням персоналізованих медичних даних для підвищення точності прогнозування цільових показників стану пацієнта і приросту інформації під час процесів опрацювання, консолідації та аналізу медичних даних. Одержані результати дають змогу:

- підвищити точність прогнозування цільових показників в підмножині простору умов на 5%, що забезпечує індивідуальний підхід до моніторингу стану пацієнта на основі тривалого спостереження та контролю лікаря;
- підвищити точність класифікації на 4 % порівняно з результатами логістичної регресії, як одного з кращих класифікаторів для набору даних по COVID - 19 та на 6 %, порівняно з результатами методу опорних векторів на поліноміальному ядрі, як кращого класифікатора для набору даних по орфанних хворобах за рахунок ієрархічної класифікації та поєднання різних моделей машинного навчання;
- підвищити точність прогнозування даних на 7-9 % та забезпечити паралелізацію процесу обробки даних як малої, так і великої розмірності за рахунок розробленої стекінгової моделі на основі Випадкового лісу як метаалгоритму та деформації метаознак під час повторного навчання на розширеному наборі даних;
- забезпечити заповнення даних за рахунок одержання додаткових значень, що керують доменом і функціональними залежностями, та включити їх у доступні навчальні дані. Правильність заповнених значень доводиться за допомогою предиктора, побудованого на вихідному наборі даних. Запропонований метод імовірнісних продукційних залежностей на 12 % покращує результати у порівнянні з Випадковим лісом (Random Forest) та Очікування-максимізація (Expectation-Maximization) для 30 % відсутніх даних;
- зменшити ймовірність появи похибки (кількість ліжкоднів) при виборі схеми лікування за рахунок персоналізації стандартних схем лікування шляхом використання простору умов та аналізу динаміки приросту значень часовозалежних даних, що забезпечує чіткість та напрямленість у пошуку рішень щодо вибору цільових схем лікування;
- розробити інформаційну систему підтримки прийняття рішень у лікуванні орфанних захворювань;
- розробити програмний модуль для підтримки прийняття рішень у визначенні схеми

лікування пацієнтів з постковідним синдромом.

Особистий внесок здобувача. У друкованих працях, опублікованих у співавторстві, ідеї та принципи, що використані в дисертаційному дослідженні, є результатом індивідуальної праці автора, а саме: [2] – запропоновано принципи опрацювання персоналізованої інформації; [3] – запропоновано інформаційну модель поведінки користувача; [4] – описано принципи опрацювання гетерогенних даних; [5] – аналіз характеристик мультимодальних даних; [6] – розроблено метод обробки малих наборів даних; [7, 17] – розроблено метод прогнозування для окремих кластерів; [8] – розроблено ієрархічний класифікатор хворих; [10, 12] – розроблено засоби прогнозування цільових змінних; [13, 14] – розроблено метод обробки персоналізованої інформації з урахуванням цільових змінних; [15] – інформаційна модель стану пацієнта; [20] – розроблено алгоритм імпутації даних; [18, 21] – принципи пошуку залежностей між цільовими змінними стану; [22] – основи класифікації індивідуальних характеристик; [16, 23] – запропоновані критерії оцінювання стану пацієнта; [24] – розроблено метод пошуку аномалій в оцінці стану хворого; [26, 27] – формалізовано процес пошуку пропусків даних; [28, 29] – запропоновані критерії ідентифікації даних; [30] – розроблено метод аналізу персоналізованої інформації; [32, 33, 34] – запропоновано персоналізований підхід до обробки та аналізу медичних даних пацієнтів; [37] – формалізовано базові операції щодо автоматизації обробки даних інформаційного об'єкта; [38] – охарактеризовано принципи застосування методів машинного навчання щодо прогнозування цільових змінних інформаційних об'єктів; [39, 40] – описано особливості використання теорії множин при опрацюванні та формалізації даних інформаційних об'єктів; [41] – окреслено особливості використання простору умов та аналізу зміни приросту значень часовозалежних даних; [42, 43] – формалізовано процес аналізу персоналізованих даних; [19, 44] – формалізовано етапи консолідації гетерогенних даних; [45] – описано особливості застосування нейронних мереж для опрацювання медичної інформації про пацієнта; [46] – описано підхід щодо опрацювання часовозалежних даних з кореляцією до термінів аналізу інформаційного об'єкта; [47, 48] – розглянуто особливості опрацювання та консолідації мультимодальних даних; [49, 50] – принципи проектування систем підтримки персоналізованих рішень; [51, 52] – спроектовано архітектуру системи управління персоналізованою інформацією; [53] – розроблено метод оптимізації аналізу персоналізованих даних інформаційного об'єкта; [54, 55] – розроблено методи опрацювання даних різного походження; [55, 56] – спроектовані етапи оптимізації опрацювання великих наборів персоналізованих даних та інформаційну систему.

Апробація результатів дисертації. Основні теоретичні та практичні результати дисертаційної роботи були представлені та обговорені на науково-практичних конференціях та семінарах: International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015, 2016, 2017; International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2016, 2017, 2018, 2020; International Conference on Advanced Computer Information Technologies, ACIT 2019, 2020; Міжнародна конференція «Інноваційні підходи і сучасна наука», м. Київ, 2015 р.; Міжнародна науково-практична конференція: «IV осінні наукові читання», м. Київ, 2015; VI Всеукраїнська науково-практична конференція «Наукові дослідження: перспективи інновацій у суспільстві і розвитку технологій», м. Харків, 2017; Міжнародна науково-практична конференція «Інформаційна безпека та інформаційні технології», ХНЕУ імені Семена Кузнеця,

2019; The 8 th International scientific and practical conference «Information, Its Impact on Social And Technical Processes», Haifa, Israel 2020; IX Міжнародна науково-практична конференція «Actual aspects of development in the context of globalization», 2020 р., Флоренція, Італія; II Міжнародна науково-практична конференція «Інформаційна безпека та інформаційні технології», 2020 р. – Кропивницький: ЦНТУ, 2020.; XVIII International Scientific and Practical Conference. Boston, USA 2020.

Публікації. Основні результати дисертаційного дослідження опубліковано у 56 наукових публікаціях, із них 21 стаття, з них: 2 статті опубліковано в журналах з Q1, 3 статті опубліковано в журналах з Q2, 7 статей у інших фахових іноземних виданнях, 7 статей у наукових фахових виданнях України, 2 монографії; 35 публікацій у матеріалах конференцій, 26 з яких індексуються в НБ Scopus. Загалом опубліковано 5 монографічних робіт, дві з яких – одноосібні, а також 3 навчальні посібники, 4 авторських свідоцтва на твір.

Структура та обсяг роботи. Дисертаційна робота складається зі вступу, шести розділів, висновків, списку використаних джерел із 302 найменувань та додатків. Обсяг дисертації становить 293 сторінки, у тому числі 239 сторінок основного тексту. Робота містить 116 рисунків та 32 таблиці.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми, сформульовано мету та основні задачі досліджень, подано короткий зміст роботи.

У першому розділі введено поняття персоналізації, персоналізованих рішень та медичних даних пацієнта. Проаналізовано існуючі практичні рішення, що відображають основні підходи до використання медичних даних та урахування персоналізації. За результатами аналізу існуючих систем штучного інтелекту у сфері медицини для вирішення проблем діагностики захворювань, дослідження геному, розробки ліків, медичної візуалізації та ін. виокремлено проблеми, які ще залишаються нерозкритими, а саме: консолідація даних про хворого, його індивідуальних особливостей, враховуючи міжнародні стандарти лікування та існуючу фармацевтичну продукцію; розроблення методів для підбору персоналізованого лікування; прогнозування результатів застосування персоналізованого лікування.

Проведено порівняльний аналіз класичних методів та визначені обмеження, які виникають при опрацюванні медичних персоналізованих даних, а саме: використання великої кількості параметрів вимагає введення евристик; мусить існувати припущення про незалежність параметрів; певні методи працюють лише з бінарними ознаками об'єктів; проблема неможливості пошуку асоціативних залежностей з малою підтримкою; наявність факту невизначеності ускладнює пошук оптимального рішення; за неоднозначності результатів виникає багато альтернативних рішень, що ускладнює процес прийняття лікарських рішень; довгий час навчання, чутливість до шуму.

За результатами досліджень сформульовані основні етапи досліджень: визначення індивідуальних особливостей, обробка персоналізованих даних, консолідація даних, класифікація загального стану, визначення важливих ознак і співвідношення їх ваги до результатів класифікації, персоналізація рішень, аналіз результатів, прогнозування наступного стану та визначення стратегії лікування чи реабілітації.

Визначені обмеження під час опрацювання універсальними методами великих та малих мультимодальних медичних наборів даних, що сформувало актуальну науково-прикладну проблему розроблення нових чи удосконалення існуючих методів штучного

інтелекту щодо підвищення точності процесу обробки медичних даних для пошуку персоналізованих рішень. Наявні методи опрацювання мультимодальних медичних даних потребують вдосконалення, оскільки у просторі умов заздалегідь невідомі зміни у стані пацієнта, що можуть супроводжуватися характеристиками даних, як викиди, шуми чи пропуски даних.

Результати першого розділу дали змогу виділити раніше невіршені проблеми та здійснити постановку задачі.

Результати розділу 1 опубліковано у [1, 2, 5, 6, 13, 14, 26, 27, 28, 29, 30, 31, 37, 51, 52].

У другому розділі введено модель відображення стану пацієнта, яка відображає його минулі та поточний стани; ця модель консолідує елементи, що представлені у вигляді взаємно-залежних та залежних від середовища оцінки множин; простір стану пацієнта подано як багатовимірну систему; формалізовано відображення фізичного стану пацієнта з урахуванням часовозалежних та часовонезалежних даних пацієнта, що дає змогу оцінити стан пацієнта в певний момент часу; уведена інформаційна модель простору станів пацієнта представлено у вигляді гіперкуба як відображення функціонального відношення загального стану пацієнта; розроблено метод пошуку шаблонів зміни стану, що базується на модифікації методу асоціативних правил.

Модель відображення стану пацієнта складається з множин значень, які є взаємозалежні та залежні від середовища оцінки.

База знань представлена як множина правил (персоналізованих рішень) R :

$$R = \{r_1, r_2, \dots\}; \quad (1)$$

Pd – множина персоналізованих даних:

$$Pd = \{p_1, p_2, \dots\}. \quad (2)$$

Для прийняття рішень використовуємо продукційні правила з множини R . При цьому встановлюється залежність між множиною персоналізованих даних та оцінкою стану пацієнта ES :

$$R: Pd \rightarrow ES \quad (3)$$

Зазначимо, що елемент множини персоналізованих рішень подано як кортеж:

$$p_i = \langle A_{in}, A_t \rangle,$$

де часовонезалежні характеристики подані як A_{in} та часовозалежні параметри пацієнта подані як A_t . Тоді ES – це множина оцінок стану пацієнта, що залежить від множини коефіцієнтів оцінки K .

Введено передумови та постумови у вигляді:

$$\langle G, A_t, Pd \rightarrow ES, K \rangle, \quad (4)$$

де імплікація $Pd \rightarrow ES$ є правило; G – передумова вибору класу правил (протокольне рішення); A_t – передумова вибору правила у класі (у нашому випадку це часовозалежні параметри, які визначають особливості стану пацієнта); K – постумова правила, що визначає перехід до наступного правила (це множина коефіцієнтів оцінки стану пацієнта, що вказують на зміни результативного показника при зміні значення вхідної змінної).

Вибір правил здійснюється на основі багатокритеріального вибору Парето, де $ES = \{e_1, e_2, \dots\}$ – векторна оцінка стану параметрів пацієнта для певних елементів множини параметрів:

$$S(ES) = \{x \in ES \mid \forall y \in ES \ [\forall i \in \{1, \dots, m\} [x_i \geq y_i]]\}. \quad (5)$$

Модель стану пацієнта, що забезпечує пошук оцінки його стану та пошук рішень щодо оптимізації процесу одужання, подана як:

$$Fe = \langle A_{in}, A_t, K, ES, S, G, R \rangle, \quad (6)$$

де Fe – система оцінки стану пацієнта, A_{in} – це множина часовонезалежних параметрів пацієнта або вхідних даних системи, що характеризують показники-фактори пацієнта, що отримують з динаміки змін його стану, які розраховуються на підставі інформації, що міститься в історії хвороби, K – це множина коефіцієнтів оцінки стану пацієнта, що вказують на зміни результативного показника при зміні значення вхідної змінної, S – множина стратегічних рішень щодо оцінки стану пацієнта, R – продукційні правила рішень щодо стратегії пошуку оптимального стану, ES – це множина оцінок стану пацієнта, що залежить від коефіцієнтів оцінки, A_t – характеристики пацієнта, що змінюються під впливом часу, G – протокольні рішення.

Для того, щоб врахувати особливості оцінки стану пацієнта, необхідно мати індивідуальний підхід до обробки його даних. Це допоможе визначити особливості пацієнта та особливості його індивідуальних характеристик. Залежності між сутностями можна відобразити бінарним відношенням.

Результати оцінки визначають: оцінку поточного стану ES , стратегічні рішення S , що визначають тактику лікування, множину показників пацієнта Pd , протокольні рішення G . Рішення щодо зміни стану пацієнта подано у вигляді n -ого відношення:

$$R = Pd \times K \times ES \times S \times G. \quad (7)$$

Прикладами рішення є перегляд та зміна тактики лікування, зміна медикаментів, зміна дозування, формування переліку лабораторних досліджень, фізіотерапевтичних заходів та ін.

Модель відображення простору станів подана як багатовимірний простір, де точка P – це його стан в певний момент часу. Отже, часовозалежні дані A_t характеризують зміни індивідуальних характеристик пацієнта під час процесу лікування, а часовонезалежні дані A_{in} визначають множину параметрів пацієнта.

Знаючи, що відображенням фізичного стану пацієнта є

$$FS(t): A_{in} \rightarrow A_t,$$

де $A_t = A_t(t)$, можемо представити з урахуванням усіх параметрів пацієнта його стан P у вигляді точки у просторі станів, де вісь x відображає часовий показник $P_o(t)$, вісь y параметричний показник $P_o(a)$, а вісь z – множину індивідів (пацієнтів) Ob . Таким чином, інформаційна модель простору станів пацієнта представлено як сукупність станів та як відображення функціонального відношення загального стану пацієнта GS .

Будь-якій трійці значень x, y, z відповідає лише одна точка у просторі $P(x, y, z)$.

У момент часу t функціональне відношення фізичного стану пацієнта формує залежність між часовим фактором та параметричним показником:

$$FS_o(t): P_o(t) \rightarrow P_o(a), \quad (8)$$

$$FS_o(t) = FS_o(A_{in}, A_t(t)). \quad (9)$$

Попередні значення фізичного стану пацієнта є визначальними для наступних станів:

$$A_{t+1}(t) = FS_o(A_{in}, A_t(t)), \quad (11)$$

$$FS_{o+1}(t) = FS_{o+1}(A_{in}, A_{t+1}(t)), A_{t+n}(t) = FS_o(A_{in}, A_{t+n-1}(t)).$$

Отже, фізичний стан пацієнта в проміжку часу подано як відношення (рисунок 1):

$$FS_{o_i} \subseteq (A_{in_i}, A_t(t)_{i_1}) \otimes (A_{in_i}, A_t(t)_{i_2}) \otimes (A_{in_i}, A_t(t)_{i_3}) \otimes \dots \otimes (A_{in_i}, A_t(t)_{i_n}). \quad (12)$$

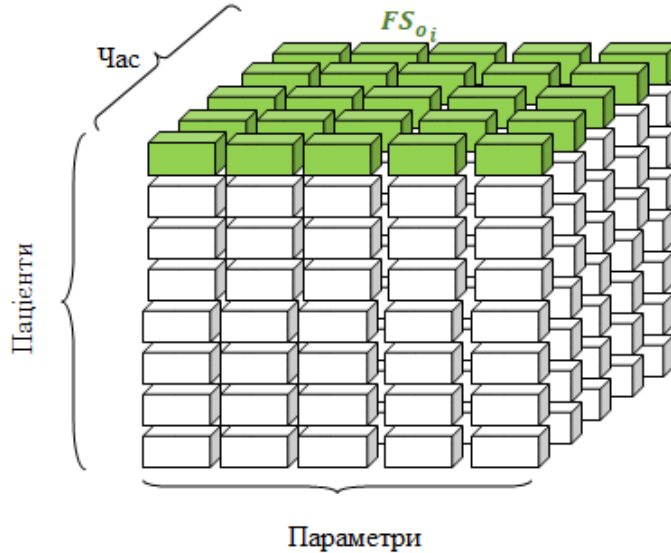


Рисунок 1. Гіперкуб параметрів фізичного стану пацієнта

Враховуючи, що FS_{oi} є показником фізичного стану пацієнта під час процесу лікування чи реабілітації, то це дає змогу проаналізувати поведінку стану пацієнта.

Отже, простір станів пацієнта з врахуванням протокольних рішень подано як:

$$PFS_{oi} \subseteq FS_{oi1}(t) \otimes FS_{oi2}(t) \otimes \dots \otimes FS_{oin}(t); \quad (13)$$

$$PFS_{oi} \subseteq FS_{oi1}(A_{ini}, A_t(t)_{i1}) \otimes FS_{oi2}(A_{ini}, A_t(t)_{i2}) \otimes \dots \otimes FS_{oin}(A_{ini}, A_t(t)_{in}),$$

$$PFS_o(t) = PFS_o(FS_o(t)), PFS_{o+1}(t) = PFS_{o+1}(FS_{o+1}(t)).$$

Відомо, що у багатьох хворих під час збору даних виникає ще множина додаткових супутніх хвороб. Кожна супутня хвороба потребує лікування з урахуванням відповідного протоколу. Тому простір станів пацієнта з врахуванням супутніх хвороб виглядає як:

$$RPFS_{oi} \subseteq PFS_{oi1}(t) \otimes PFS_{oi2}(t) \otimes \dots \otimes PFS_{oim}(t); \quad (14)$$

$$RPFS_{oi} \subseteq PFS_{oi1}(FS_{o1}(t)) \otimes PFS_{oi2}(FS_{o2}(t)) \otimes \dots \otimes PFS_{oim}(FS_{oin}(t)),$$

$$RPFS_o(t) = RPFS_o(PFS_o(t)), RPFS_{o+1}(t) = RPFS_{o+1}(PFS_{o+1}(t)).$$

Тоді результуючий простір стану пацієнта GS_{oi} подано як перетин усіх просторів станів пацієнта $RPFS_o(t)$, $PFS_o(t)$, $FS_o(t)$ в певний момент часу:

$$GS_{oi} = RPFS_o(t) \cap PFS_o(t) \cap FS_o(t). \quad (15)$$

У розділі розроблено метод пошуку зміни стану пацієнта, який базується на модифікації алгоритму пошуку асоціативних правил. Цей метод є удосконаленням методу упорядкованого пошуку рішень з вибору цільових схем лікування, і дає змогу зменшити імовірність появи похибки при виборі схеми лікування. Як міра статистичної значущості використано показник Кульбака-Лейблера.

Існуючі алгоритми пошуку асоціативних правил дають змогу шукати лише залежності, визначені загальним набором вхідних даних; крім того, вони мають високу обчислювальну складність, якщо існує багато правил класифікації. Водночас, для великих наборів даних, з наперед невідомою структурою, асоціації з високою підтримкою практично відсутні; зазвичай рівень підтримки знайдених асоціацій є низьким. Проте, зважаючи на особливості існуючих алгоритмів пошуку асоціативних правил, асоціації, що не відповідають мінімальному порозу підтримки, можуть бути виключені з розгляду, хоча можуть представляти інтерес.

Для вирішення цієї проблеми в дисертації запропоновано шукати асоціативні правила не лише для окремих об'єктів та їх значень, а й для ієрархії об'єктів. Якщо на

нижчих ієрархічних рівнях немає цікавих асоціацій, то вони можуть виникати на вищих рівнях. Іншими словами, підтримка окремого об'єкта завжди буде меншою, ніж підтримка групи, до якої він належить:

$$S(I) > S(i_j), \quad (15)$$

де I – група знаходиться в ієрархії; i_j – елемент, включений до даної групи. Причини цього очевидні: загальна підтримка групи дорівнює сумі підтримки для включених до неї предметів:

$$S(I) = \sum_{j=1}^n i_j, \quad (16)$$

де n – кількість елементів у групі.

Асоціативні правила, знайдені для об'єктів чи подій, розташованих на різних ієрархічних рівнях, називаються багаторівневими правилами. Спускаючись до нижчих рівнів абстракції, аналізуються елементи лише тих категорій та підкатегорій, які є частими наборами, тобто зустрічаються принаймні задалегідь визначену кількість разів, де k – номер рівня.

Треступеневий метод пошуку зміни стану пацієнта:

1. Кластеризація пацієнтів методом k середніх із попереднім визначенням кількості кластерів методом ліктя використовується для пошуку поведінки стану;
2. Побудова ієрархії шаблонів у кожному кластері на основі модифікованого алгоритму Аргіогі та міри Кульбака-Лейблера – для пошуку послідовності зміни стану;
3. Пошук найчастішого стану в кластері на різних рівнях ієрархії – для передбачення наступного стану хворого. Для кожного кластера:

- 3.1. Сортуння за кодом пацієнта, далі – за датою та схемою лікування. *Результат* – послідовність зміни стану пацієнтів в кластері:

$$T_D = \{t_{id1}, t_{id2}, \dots, t_{idm}\} \quad (17)$$

Вхід: послідовність транзакцій вихідної бази $\langle t_{id1}, t_{id2}, \dots, t_{idm} \rangle$,

Вихід: перестановка

$$\langle t'_{id1}, t'_{id2}, \dots, t'_{idm} \rangle,$$

Для вихідної послідовності, для усіх її членів виконується співвідношення:

$$t'_{id1} \leq t'_{id2} \leq \dots \leq t'_{idm}.$$

- 3.2 Пошук частих станів FS' та їх характеристик:

$$FS' = \{p \in P | X_i = X_j, Conf(X \rightarrow Y) \geq MinConf\}. \quad (18)$$

- 3.3 Пошук послідовностей з частими станами:

$$Lift(X \rightarrow Y) = \frac{Supp(X \rightarrow Y)}{Supp(X) * Supp(Y)}, \quad (19)$$

$$Lev(X \rightarrow Y) = Supp(X \rightarrow Y) - Supp(X) * Supp(Y).$$

- 3.4 Пошук частих послідовностей.

- 3.5 Серед знайдених частих послідовностей пошук довгих послідовностей.

Далі у розділі розроблено *метод консолідації мультимодальних даних*, який базується на оцінюванні їх якості та корисності:

Етап 1. Визначення списку параметрів, необхідних для формування простору станів GS_{oi} ;

Етап 2. Визначення джерела даних для кожного параметру з GS_{oi} . Якщо коефіцієнт оцінки k більший за порогове значення, провести екстракцію значень;

Етап 3. Пошук відхилень даних від нормованих показників по кожному типу захворювання;

Етап 4. Формування результуючого набору даних.

Особливістю розробленого методу консолідації є попереднє визначення структур даних (етап 2) та узгодження семантики (етап 3).

Результати розділу 2 опубліковано у [1, 2, 3, 5, 9, 11, 13, 14, 15, 20, 21, 23, 24, 25, 26, 28, 29, 30, 32, 35, 38, 39, 44, 46, 48, 49, 53].

У **третьому розділі** розроблено метод двоетапної обробки малих наборів даних, що базується на основі ієрархічного предиктора для підвищення точності процесу прогнозування та узагальнення результатів; розроблено новий метод ансамблювання моделей машинного навчання на основі модифікації стекінгу для збільшення повторюваності результатів на малих вибірках даних.

Малі набори даних характеризуються невеликою кількістю екземплярів (десятки чи сотні) та великою кількістю параметрів (також десятки чи сотні). Аналізуючи малі набори даних, виникає ряд проблем, які суттєво впливають на результати класифікації/прогнозування: ймовірність появи нових значень у тестовому наборі даних, і яких не було у тренувальному, зміна розподілу на тестовому наборі даних порівняно з тренувальним, різна вага параметрів для тренувального та тестового наборів, перенавчання моделі. Водночас, малі набори даних є характерними для галузі медицини, зокрема це стосується інформації про пацієнтів з орфанними хворобами.

Для зменшення впливу вище перелічених проблем у розділі розроблено *метод двоетапної обробки малих даних* з використанням ієрархічного предиктора, який передбачає два етапи аналізу даних:

1. Кластеризація:

1.1. За допомогою статистики розривів шукається кількість кластерів;

1.2. Застосування алгоритму k середніх для знайденої кількості кластерів:

$$\min \left[\sum_{j=1}^k \sum_{i=1}^N \|x_i - cx_j\|^2 \right], \quad Cost = \sum_{i=1}^k \sum_{j=1}^{S_i} d_{ij} d_s(x_j, cx_i), \quad (20)$$

де n – кількість об'єктів, x – значення показника, cx – центроїд.

2. Побудова моделі прогнозування/класифікації для кожного кластера:

2.1. Пошук слабких предикторів;

2.2. Узагальнення результатів за допомогою жорсткого голосування:

$$\begin{cases} \sum_j w_{i,j} x_{i,j,n} = y_{i,n} \\ \sum_j w_{i,j} = 1 \end{cases} \quad (21)$$

де w – вага оцінки слабого предиктора, y – результат прогнозування.

Ієрархічний предиктор застосовувався на наборі даних зі 264 екземплярів для визначення параметрів пацієнта, що впливають на резистентність до захворювання на COVID - 19. Розв'язано задачу бінарної класифікації. В якості слабких предикторів використано логістичну регресію, опорні вектори з поліноміальним ядром, наївний Байєсовий класифікатор, XGBoost, Випадковий ліс на 500 дерев з максимальною глибиною 8, Нейронна мережа з одним прихованим шаром 5-ма нейронами та дерево підтримки прийняття рішень, побудоване за алгоритмом CART. Гіперпараметри моделей підбиралися з допомогою GridSearch. Слабкі предиктори застосовувалися також на різних наборах ознак (таблиця 1). Метод відбору важливих ознак описано в розділі 4.

Таблиця 1. Точність моделей для обраних ознак

Модель	Вік, перенесений грип, IgG, IgM	Група_крові, Вік, Стать, Група_крові, перенесений грип
Логістична регресія	0,633	0,671
Опорні вектори	0,671	0,722
Наївний Байес	0,674	0,732
XGBoost	0,935	0,945
Випадковий ліс	0,945	0,934
Нейронна мережа	0,832	0,845
Дерево рішень	0,553	0,631

Точність ієрархічного предиктора наведена в таблиці 2.

Таблиця 2. Точність ієрархічного предиктора

Модель	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Ієрархічний класифікатор	0,941	0,945	0,961	0,957

Ієрархічний предиктор демонструє найкращу точність для кластерів 3 і 4. Лише точність по кластеру 1 домінується слабким предиктором Random Forest на цілому наборі даних. XGBoost і Випадковий ліс алгоритми також дають високу точність для моделі, заснованої на обраних ознаках, але меншу в порівнянні з розробленим предиктором.

У розділі розроблено *метод групування моделей машинного навчання*, як стекінгову модель на основі алгоритмів машинного навчання для забезпечення точності прогнозування даних, що базується на використанні Випадкового лісу, як метаалгоритм, базується на деформації метаознак і повторному навчанні на розширеному наборі даних, що забезпечує підвищення точності прогнозування даних та паралельної обробки даних як малої, так і великої розмірності.

Математичне формулювання розробленої моделі наступне:

1. Формування k -перехресних згорток, випадково згенерованих із початкового набору даних:

$$\{z_1^1, \dots, z_B^1\}, \{z_1^2, \dots, z_B^2\}, \dots, \{z_1^K, \dots, z_B^K\}, \quad (22)$$

де K – кількість піднаборів, B – розмір піднабору, z_b^l – спостереження l -го зразка.

2. Навчання K незалежних слабких предикторів

$$w_1(\cdot), w_2(\cdot), \dots, w_K(\cdot)$$

та поєднання результатів навчання за допомогою метамоделі mw :

$$s_K(\cdot) = mw(w_1(\cdot) \times w_2(\cdot), w_1(\cdot) \times w_3(\cdot), \dots, w_{K-1}(\cdot) \times w_K(\cdot)), \quad (23)$$

де $w_i(\cdot) \times w_j(\cdot)$ – результат попарного множення слабких предикторів.

3. Використання спотворених функцій разом із навчальним набором даних у метамоделі. Ця комбінація дозволяє уникнути кореляції результатів слабких предикторів і збільшує узагальнення моделі.

Основним недоліком класичної стекінгової моделі S_k є те, що метаатрибути тренування та тестування різні на малих наборах даних. Метаатрибут у навчальній вибірці не є відповідями конкретного регресора, він складається з частин, які є результатами різних регресій (з різними коефіцієнтами). Метаатрибут на контрольній вибірці, взагалі, є відповіддю на зовсім іншу регресію, налаштовану на повне навчання. Розроблена модель стекінгу деформує метаознаки за результатами попарного множення. Далі, спотворені функції використовуються разом із навчальним набором даних у метамоделі. Ця комбінація дозволяє уникнути кореляції результатів слабких предикторів і збільшує узагальнення моделі.

Результати розділу 3 опубліковано у [1, 4, 6, 7, 8, 10, 15, 17, 18, 20, 27, 28, 30, 32, 40, 41, 42, 43, 50, 51, 54]

У **четвертому розділі** проаналізовано існуючі методи відбору важливих ознак: фільтри, вбудовані алгоритми. Розроблено гібридний ансамбль вибору ознак на основі комбінації селекторів та агрегування результатів, який використовується на етапі попередньої обробки на великих наборах даних.

Відбір важливих ознак дає змогу підвищити точність моделей класифікації/прогнозування, а також зменшити обчислювальну складність при опрацюванні великих наборів даних. Вбудовані алгоритми виконують вибір ознак під час процедури навчання класифікатора, і вони оптимізують набір ознак, що використовуються для досягнення вищої точності. Основна перевага вбудованих алгоритмів у тому, що вони зазвичай знаходять рішення швидше, уникаючи повторного навчання даних з нуля, усуваючи необхідність розділяти дані на навчальну та тестову підвибірки. Однак ці алгоритми не є універсальними.

У розділі розроблено *метод зменшення розмірності вхідних даних*, який за рахунок класифікаторів, асоціативних правил та узагальненого рангу ознак забезпечує підвищення точності вибору пріоритетних ознак на великих наборах даних. Це передбачає тришарове стекування ансамблю моделей методів, що дає змогу об'єднати асоціативну класифікацію зі слабкими класифікаторами в ансамбль для узагальнення результатів.

Основні етапи наведено нижче:

1. Пошук асоціативних правил для пошуку прихованих залежностей даних;
2. Застосування кількох селекторів (Ворута, Випадковий ліс, Дерево підтримки прийняття рішень);
3. На основі кореляційної матриці відкидаються ознаки, які корелюються;
4. Узагальнення рейтингу ознак на основі індексу Жакара:

$$(S_1, \dots, S_{1n}) = \frac{|S_1 \cap \dots \cap S_n|}{|S_1 \cup \dots \cup S_n|} \quad (24)$$

де S_i – підмножина ознак на i -й ітерації, $i=1, \dots, n$.

Схема вибору ознак гібридного ансамблю наведена на **рисунку 2**.

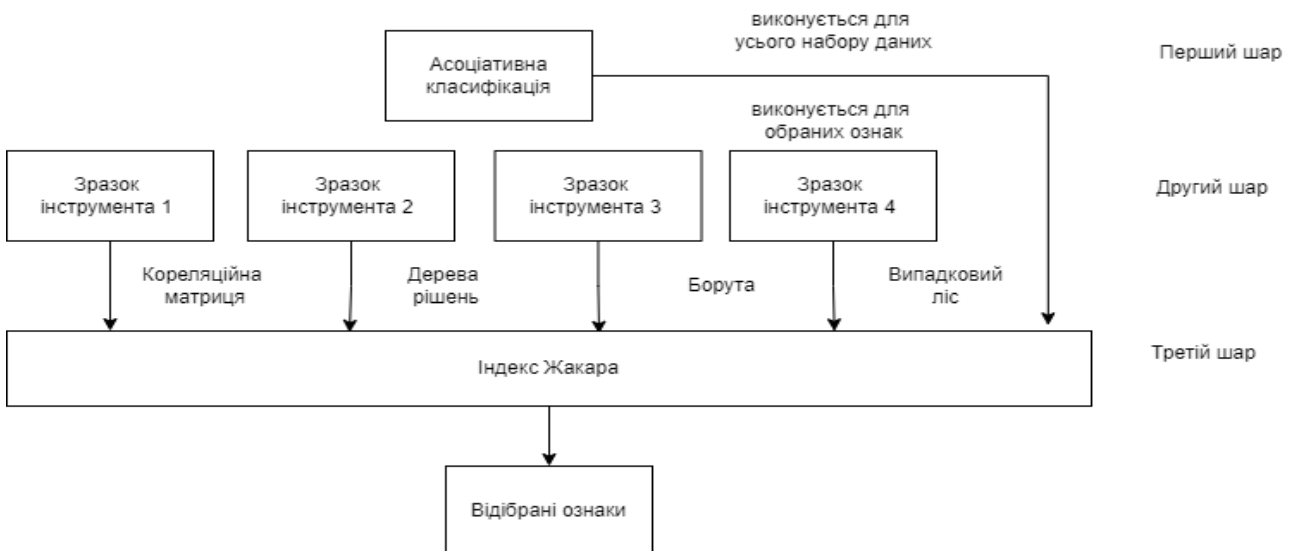


Рисунок 2. Схема вибору ознак гібридного ансамблю.

Розроблений метод відбору ознак тестувався на задачі прогнозування наявності постковідного ефекту. Задача зведена до задачі бінарної класифікації: короткий

постковід «Short» та довгий постковід «Long» (більше 21 дня). Набір даних збалансований. У таблиці 4 наведено точність класифікації для п'яти класифікаторів усього набору даних та вибраних ознак. Кожен з класифікаторів на наборі з вибраних ознак показує кращий результат по якісних метриках точності.

Таблиця 4. Точність класифікації для набору ознак, отриманих на основі моделі гібридного ансамблю

Модель	AUC		Accuracy		F1		Точність		Відкликання	
	цілий	вibr.	цілий	вibr.	цілий	вibr.	цілий	вibr.	цілий	вibr.
Дерево прийняття рішень	0,782	0,792	0,694	0,720	0,724	0,750	0,788	0,808	0,700	0,730
Машина Опорних векторів	0,901	0,922	0,845	0,860	0,831	0,846	0,901	0,958	0,789	0,821
Наївний Байєс	0,865	0,883	0,845	0,860	0,854	0,877	0,901	0,926	0,822	0,833
Відкалібрований учень	0,867	0,867	0,860	0,860	0,868	0,868	0,999	0,999	0,767	0,767
Логістична регресія	0,834	0,867	0,801	0,840	0,812	0,846	0,921	0,999	0,733	0,752

Результати розділу 4 опубліковано у [1, 5, 6, 9, 10, 11, 12, 17, 19, 21, 24, 28, 29, 32, 43, 44, 54, 55].

У п'ятому розділі визначені типи аномалій та шляхи їхнього пошуку, що дає можливість диференціації підходів щодо аналізу персоналізованих медичних даних та визначені особливості їхнього застосування; визначені основні проблеми, що виникають при обробці медичних персоналізованих даних, за рахунок чого сформовано задачу розроблення ефективної стратегії аналізу даних; розроблено метод заповнення відсутніх даних на основі ймовірнісних продукційних залежностей, що збільшує стійкість моделі до помилок у даних та забезпечення аналізу мультимодальних даних при паралельній реалізації в розподілених базах даних.

Виявлення аномалій – це пошук непередбачуваних, рідкісних значень, подій або шаблонів послідовностей в потокових або табличних даних. Особливістю медичних даних є залежність аномалій від часу. Коли в моніторингових системах метрик стає надто багато, то відслідковування аномалій ускладнюється або навіть стає неможливим. У таких випадках зазвичай застосовують перевірку на досягнення метрикою певного критичного значення, проте не для всіх типів даних така перевірка є відповідною. Аномалії бувають різних типів, а саме: викиди, зсув, відхилення патерну, зміна характеру розподілу та спільні аномалії.

Послідовностей даних може бути більше ніж одна, вони можуть залежати одна від одної. Так, поведінка однієї послідовності може визначати як поведе себе інша. Без цієї інформації неможливо точно сказати, чи це аномалія чи ні, а для цього система має приймати на вхід не одну, а множину послідовностей та знаходити часові закономірності між ними. Методи аналізу даних ефективно використовуються на чітких і попередньо оброблених даних. Тому у розділі розроблено метод попередньої обробки даних та заповнення пропусків даних.

Ймовірнісна продукційна залежність (Probability productive dependency, PPD) – це залежність, схожа на асоціативне правило у виборі первинного співвідношення, що є правильним для багатьох пацієнтів. Поріг значущості повинен бути визначений експертно або на основі оцінки ймовірності помилкового вибору цієї залежності.

Основна відмінність між асоціативними правилами та PPD полягає в тому, що PPD створено з існуючих функціональних залежностей (FD) у наборі даних.

$$F_I: K = \{a_i\}, a_i \in A, D = \{a_j\}, a_j \in A, : P(k \in K \rightarrow d \in D) = p, \quad (25)$$

де k і d – кортежі груп атрибутів K і D відповідно.

Метод заповнення пропусків даних на основі ймовірнісних продукційних залежностей збільшує стійкість моделі до помилок в даних та забезпечення аналізу мультимодальних даних при паралельній реалізації в розподілених базах даних, що ґрунтується на застосуванні двох алгоритмів: видобування PPD та імпутації відсутніх медичних персоналізованих даних.

Розроблено алгоритм видобування PPD:

1. Множина правил PPD-set порожня;
2. Згрупувати сутності з однаковими X -значеннями;
3. Вибрати атрибути з FD з тим самим X і додати їх до Y ;
4. Обчислити підтримку та достовірність по кортежах, вибраних на кроці 2;
5. Визначити кортежі з найвищим значенням достовірності;
6. Додати $X \rightarrow Y$ до PPD-set.

Алгоритм заповнення даних:

1. Повнота = 0;
2. Поки Повнота / 100 < Імпутація;
3. Упорядкуйте всі атрибути з PPD-set за рівнем надійності;
4. Для кожної групи:
 - 4.1. Якщо відсоток непорожній значення Y вище або дорівнює Підтримці слід заповнити порожні значення за допомогою PPD-set:
 - 4.1.1. Повнота++;
 - 4.2. Інакше:
 - 4.2.1. Об'єднати PPD за допомогою правил Армстронга.

Розроблений метод порівняно з існуючими методами імпутації: асоціативними правилами (AR), Випадковим лісом (RF), Машиною опорних векторів (SVM), багат шаровим перцептроном (MLP), Очікуванням - Максимізацією (EM) і k - найближчим сусідом (KNN) (рисунок 3). Помилка відновлення представлена за допомогою нормалізованої середньоквадратичної помилки (NRMSE).

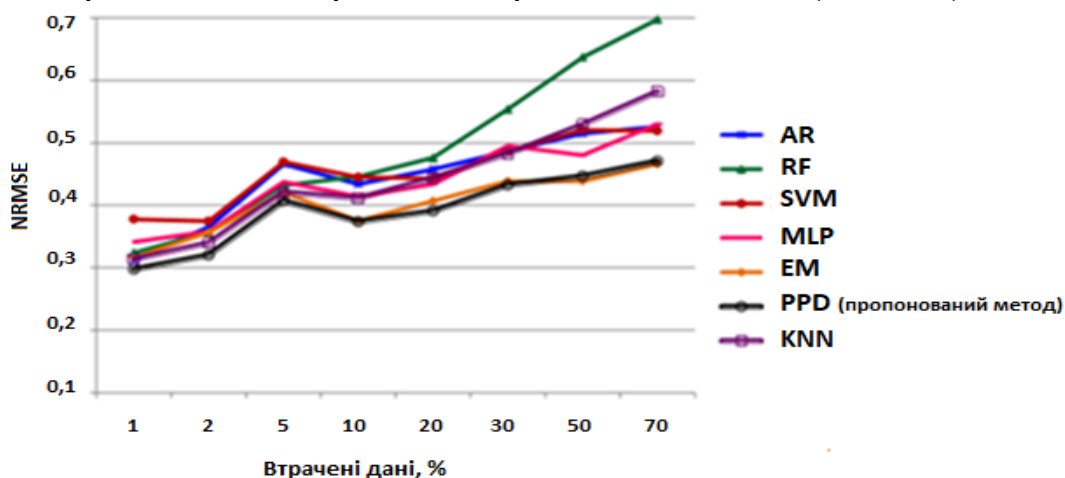


Рисунок 3. Помилка відновлення NRMSE

PPD відновлює пропущені дані на 12% краще, ніж моделі RF та EM при 30 % відсутніх даних. Метод EM виглядає найкращим для додаткових відсутніх даних (40 – 50 % відсутніх даних), а PPD дає еквівалентні результати з SVM. Результати

порівняння підтвердили переваги розробленого методу:

- збереження характеристик стійкості до помилок у даних;
- можливість паралельної реалізації в розподілених базах даних;
- автоматизація та виконання аналізу різних типів даних.

Результати розділу 5 опубліковано у [3, 5, 12, 14, 16, 17, 20, 22, 27, 29, 30, 39, 41, 56].

У шостому розділі розроблено архітектуру системи підтримки прийняття медичних рішень щодо прогнозування станів пацієнта на підставі опрацювання та аналізу персоналізованих медичних даних; подано опис імплементації рішень у системі підтримки прийняття рішень для лікування хворих на імунодефіцити (ПІД) з порушеннями антитілоутвореннями; подано результати розробки програмного модулю щодо прогнозування динаміки поширення COVID - 19; подано результати імплементації системи для виявлення помилок, що дає змогу експериментувати із різними параметрами для моделі, знову і знову на тих самих даних.

Розроблена архітектура системи підтримки прийняття медичних рішень щодо прогнозування станів пацієнта на підставі опрацювання та аналізу персоналізованих медичних даних складається з таких компонентів (рисунок 4).

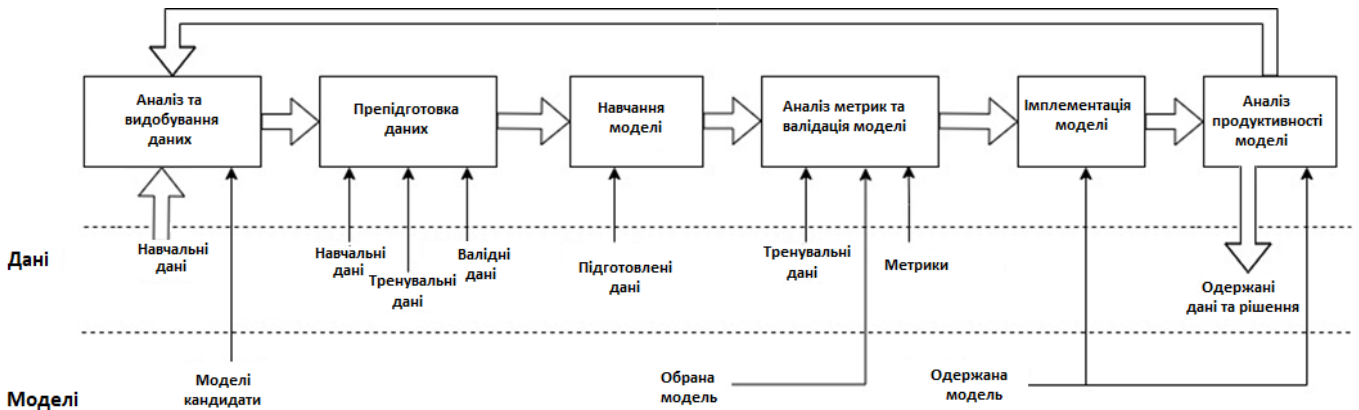


Рисунок 4. Схема процесу підтримки прийняття персоналізованих рішень

На рисунку 5 зображена структура розробленої інформаційної системи підтримки прийняття рішень.

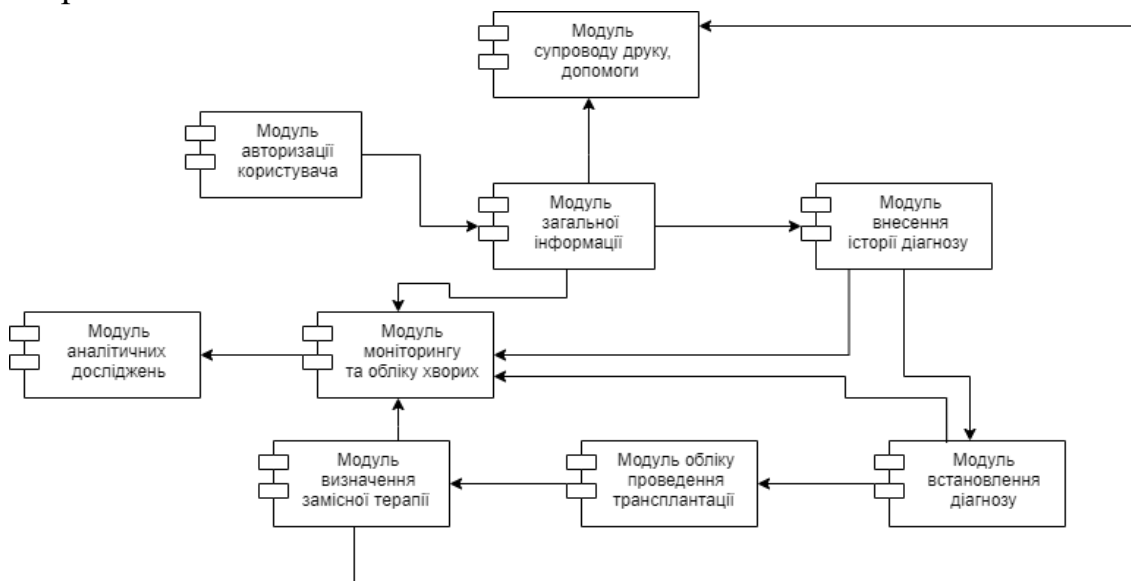


Рисунок 5. Структура інформаційної системи підтримки прийняття рішень

Наступною імплементацією є система прогнозування тривалості захворювання,

хворих на COVID - 19 (прогнозувалася кількість ліжокднів, ознаками були параметри пацієнтів та схема лікування), а також прогнозування тривалості постковідного періоду. У системі імплементовані результати застосування на основі ансамблю моделей машинного навчання для вибору пріоритетних ознак на великих наборах даних, який складається з класифікаторів, асоціативних правил та узагальненого рангу ознак на основі індексу Жакара, що дозволяє уникнути кореляції ознак та збільшує узагальнення моделі.

Імплементовано програмний модуль щодо виявлення аномалій за значеннями параметрів пацієнтів.

Результати розділу 6 опубліковано у [2, 5, 7, 8, 15, 16, 17, 18, 22, 33, 34, 36, 45, 47, 55, 57]

У додатку наведені акти впровадження результатів дисертаційної роботи.

ВИСНОВКИ

За результатами проведеного дослідження вирішено важливу науково-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізація результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта. У результаті виконання цієї роботи одержані наступні результати:

1. Здійснено аналіз процесів опрацювання мультимодальних персоналізованих медичних даних, попередньої обробки даних та аналізу якості моделі даних, що дозволило визначити проблему та задачі дослідження;
2. Обґрунтовано актуальність розв'язання цієї проблеми на основі введення модель відображення стану пацієнта в багатовимірному просторі умов, яка за рахунок додаткових вимірів та параметрів забезпечила підвищення точності прогнозування цільових змінних та дала змогу підвищити точність прогнозування цільових показників в підмножині простору умов на 5 %, що забезпечило індивідуальний підхід до моніторингу стану пацієнта на основі тривалого спостереження та контролю лікаря;
3. Вперше розроблено метод двоетапної обробки даних на основі ієрархічного предиктора, що забезпечує підвищення точності процесу узагальнення результатів на малих наборах даних у порівнянні з результатами логістичної регресії, як кращого класифікатора для набору даних по COVID - 19 та результатами поліноміального методу опорних векторів, як кращого класифікатора для набору даних по орфанних хворобах за рахунок ієрархічної класифікації та поєднання різних моделей машинного навчання;
4. Вперше розроблено метод групування моделей машинного навчання як стекінгову модель для забезпечення точності прогнозування даних на 7-9 % та паралелізації процесу обробки даних, що на відміну від існуючих моделей убезпечує універсальність пошуку рішень щодо малих та великих наборів персоналізованих даних;
5. Вперше розроблено метод зменшення розмірності вхідних даних на основі ансамблю моделей слабких предикторів для вибору найважливіших ознак, що базується на прирості інформації з урахуванням результатів застосування декількох селекторів та агрегації кінцевих результатів на підставі урахування індекса Жакара для оцінки подібності одержаних підмножин ознак та проведення мажоритарного

- голосування результатів, що дає змогу уникнути кореляції результатів слабких предикторів і збільшує узагальнення моделі;
6. Удосконалено метод заповнення пропусків даних, який ґрунтується на пошуку подібності даних, що можуть впливати на доменні значення та функціональні залежності між ними і забезпечує їхнє включення у навчальні дані, що дає змогу забезпечити стійкість до помилок даних, паралелізації обчислень та аналізу різнотипних даних. Розроблений метод заповнення даних на 12% покращує результати на відміну від моделі Випадкового лісу (Random Forest) та Очікування - Максимізацією (Expectation-Maximization) для 30% відсутніх даних;
 7. Удосконалено метод пошуку зміни стану пацієнта шляхом використання простору умов та аналізу зміни приросту значень часовозалежних даних, що на відміну від методу упорядкованого пошуку надає чіткості та направленості у пошуку рішень стосовно вибору цільових схем лікування, що дає змогу зменшити імовірність появи похибки (кількість ліжкоднів) при виборі схеми лікування, за рахунок агрегованого опрацювання даних з різних просторів моделі;
 8. Удосконалено метод консолідації мультимодальних даних за рахунок попереднього визначення структур даних та узгодження семантики, що на відміну від методів консолідації даних, на рівні сховища даних, дало змогу забезпечити агрегування даних з різною структурою;
 9. Розроблені методи використані для розроблення програмних модулів інформаційних систем підтримки прийняття рішень у лікуванні орфанних хвороб та для підтримки прийняття рішень щодо адаптації схеми лікування пацієнтів з постковідним ефектом;
 10. Практичне значення результатів дисертаційного дослідження полягає у підвищенні точності прогнозування цільових показників у підмножині простору умов на 5 %, підвищенні точності класифікації на 4 % порівняно з результатами логістичної регресії як кращого класифікатора у порівнянні з існуючими для набору даних по COVID – 19, підвищенні точності прогнозування даних на 7-9% та забезпеченні паралелізації процесу обробки даних як малої, так і великої розмірності, забезпеченні заповнення пропусків даних за рахунок одержання додаткових значень, а також зменшенні ймовірності появи похибки (кількість ліжкоднів) при пошуку шаблонів динаміки стану пацієнта шляхом використання простору умов;
 11. Достовірність наукових та практичних результатів підтверджується відповідними матеріалами про впровадження дисертаційних досліджень, а також за рахунок порівняння одержаних практичних результатів з результатами застосування існуючих класичних методів та підходів щодо опрацювання персоналізованої інформації про пацієнта.

СПИСОК ОСНОВНИХ ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Melnykova, Nataliia, et al. 'Data-Driven Analytics for Personalized Medical Decision Making'. *Mathematics*, vol. 8, no. 8, July 2020, p. 1211, <https://doi.org/10.3390/math8081211>. (індексована в наукометричній базі Scopus та Web of Science Core Collection, квартал Q1 відповідно до класифікації SCImago Journal);
2. Melnykova, Nataliia, et al. 'Personalized Data Analysis Approach for Assessing Necessary Hospital Bed-Days Built on Condition Space and Hierarchical Predictor'. *Big Data and Cognitive Computing*, vol. 5, no. 3, Aug. 2021, p. 37,

- <https://doi.org/10.3390/bdcc5030037>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*, квартал Q1 відповідно до класифікації *SCImago Journal*);
3. Melnykova, Nataliia, et al. 'The Hierarchical Classifier for COVID-19 Resistance Evaluation'. *Data*, vol. 6, no. 1, Jan. 2021, p. 6., <https://doi.org/10.3390/data6010006>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*, квартал Q2 відповідно до класифікації *SCImago Journal*);
 4. Melnykova, Nataliia, et al. 'An Ensemble Methods for Medical Insurance Costs Prediction Task'. *Computers, Materials and Continua*, vol. 70, no. 2, 2022, pp. 3969 – 84, <https://doi.org/10.32604/cmc.2022.019882>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*, квартал Q2 відповідно до класифікації *SCImago Journal*);
 5. Melnykova, Nataliia, et al. 'The Ensembles of Machine Learning Methods for Survival Predicting after Kidney Transplantation'. *Applied Sciences*, vol. 11, no. 21, Nov. 2021, p. 10380, <https://doi.org/10.3390/app112110380>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*, квартал Q2 відповідно до класифікації *SCImago Journal*);
 6. Мельникова, Н. І. 'Особливості опрацювання медичної інформації для систем підтримки прийняття лікувальних рішень'. *Вісник Національного університету Львівська політехніка. Інформаційні системи та мережі*, no. 832, pp.190 – 204;
 7. Шаховська Н. Б., Мельникова Н. І. 'Методи побудови моделі поведінки користувачів'. *Український журнал інформаційних технологій*, 2020, no. 2, vol. 1, pp. 43–51;
 8. Мельникова Н.І., та ін. 'Розроблення інформаційної технології опрацювання персоналізованих медичних даних'. *Вісник Національного університету "Львівська політехніка". Інформаційні системи та мережі: збірник наукових праць*, 2015, no. 814, pp. 90 – 99.;
 9. Shakhovska, N. B., and Melnykova. N. I. 'Нові Методи Та Рішення Щодо Побудови Моделі Поведінки Користувачів'. *Scientific Bulletin of UNFU*, vol. 30, no. 5, Nov. 2020, pp. 76 – 83., <https://doi.org/10.36930/40300513>.;
 10. Кривенчук Ю. П., Шаховська Н. Б., Вовк О. Б., Мельникова Н. І. 'Комп'ютерне моделювання функцій перетворення оптичних схем засобу вимірювання температури, побудованого на ефекті рамана та структура алгоритму їх дослідження'. *Радіоелектроніка, інформатика, управління*, 2018, no. 346, pp. 25 – 33. DOI:10.15588/1607-3274-2018-3-3. (індексована в наукометричній базі *Web of Science Core Collection*);
 11. Melnykova Nataliia, Basystiuk Oleh. 'Multimodal Speech Recognition Based on Audio and Text Data'. *Вісник ХНУ. Серія: технічні науки*, 2022, no 3.;
 12. Мельникова Н. І., Поберейко П. Б. 'Дослідження методів пошуку ключових кадрів у відеопотоці з використанням нейронних мереж для систем пошуку' *Вісник Хмельницького національного університету. Серія: Технічні науки*, 2022. vol. 3, no. 309. pp. 55 – 61.;
 13. Melnykova, Nataliia, et al. Specifics personalized approach in the analysis of medical information. *Econtechmod: an international quarterly journal on economics in technology, new technologies and modelling processes*. Lublin. Rzeszow, 2016. vol. 5, no. 2. pp. 109 – 116.;

14. Kryvenchuk, Y., Shakhovska, N., Melnykova, N., and Boichuk, A. 'Organization of the network connection in the Industry 4.0'. *Econtechmod: An International Quarterly Journal on Economics of Technology and Modelling Processes*, 8, 2019, pp. 39 – 45;
15. Melnykova N. 'Using of personalized approach for assessment of the financial condition of the company' *Econtechmod: an international quarterly journal on economics in technology, new technologies and modelling processes*. Lublin, Rzeszow, 2017, vol. 6, no 2, pp. 39 – 44;
16. Melnykova N. 'Analysis of the Data Mining and Classification of Patients' States'. *Manažérska Informatika*, 2020, no. 2 <https://manazerskainformatika.sk/analysis-of-the-data-mining-and-classification-of-patients-states/>;
17. Melnykova N. 'Method for Clustering and Determining the Average Distance between Clusters'. *Manažérska Informatika*, 2021, no. 2, <https://manazerskainformatika.sk/method-for-clustering-and-determining-the-average-distance-between-clusters/>;
18. Melnykova N. 'A New Approach to Modelling the Nature of Individual Morbidity Using Partial Functional Dependencies'. *Manažérska Informatika*, 2021, no. 1, <https://manazerskainformatika.sk/a-new-approach-to-modelling-the-nature-of-individual-morbidity-using-partial-functional-dependencies/>;
19. Melnykova N. 'Model of States Warehouse of a State of the State of the Object and Personalized Decisions'. *Manažérska Informatika*, 2022, no. 1, <https://manazerskainformatika.sk/model-of-states-warehouse-of-a-state-of-the-state-of-the-object-and-personalized-decisions/>;
20. Мельникова Н. І., and Мельников В. А. 'Персоналізований підхід до обробки та аналізу медичних даних пацієнтів'. *Інформаційна безпека та інформаційні технології : монографія*, Харків : Вид. Рожко С. Г., 2019., pp. 247 – 259;
21. Мельникова Н.І. 'Методи оптимізації рішень щодо аналізу персоналізованих даних'. *Кібербезпека та інформаційні технології: монографія*. – Х.: ТОВ «ДІСА ПЛЮС», 2020, ISBN 978-617-7927-01-2, pp. 210 – 225;
22. Melnykova, Nataliia, et al. 'Big Data Analysis in Development of Personalized Medical System'. *Procedia Computer Science*, vol. 160, 2019, pp. 229 – 34, <https://doi.org/10.1016/j.procs.2019.09.461>. *(індексована в наукометричній базі Scopus та Web of Science Core Collection)*;
23. Melnykova, Nataliia, et al. 'Using Big Data for Formalization the Patient's Personalized Data'. *Procedia Computer Science*, vol. 155, 2019, pp. 624 – 29, <https://doi.org/10.1016/j.procs.2019.08.088>. *(індексована в наукометричній базі Scopus та Web of Science Core Collection)*;
24. Melnykova, Nataliia, et al. 'Anomalies Detecting in Medical Metrics Using Machine Learning Tools'. *Procedia Computer Science*, vol. 198, 2022, pp. 718 – 23, <https://doi.org/10.1016/j.procs.2021.12.312>. *(індексована в наукометричній базі Scopus та Web of Science Core Collection)*;
25. Melnykova, Nataliia. 'A Novel Approach for the Automatic Detection of COVID in a Patient by Using a Categorization Methods'. *Procedia Computer Science*, vol. 198, 2022, pp. 712 – 17, <https://doi.org/10.1016/j.procs.2021.12.311>. *(індексована в наукометричній базі Scopus та Web of Science Core Collection)*;
26. Melnykova, Nataliia. 'Model of the system of personalized analysis of financial condition of the enterprise'. *Advances in Intelligent Systems and Computing*, 2018, no. 689, pp. 334-345, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85036467510&>

- [doi=10.1007%2f978-3-319-70581-1_24&partnerID=40&md5=909d5e2a66315b464c19399e7482a86e](https://doi.org/10.1007%2f978-3-319-70581-1_24&partnerID=40&md5=909d5e2a66315b464c19399e7482a86e). (індексована в наукометричній базі Scopus, кuartиль Q3 відповідно до класифікації SCImago Journal);
27. Melnykova, Nataliia., et. al. ‘Smart Integrated Robotics System for SMEs Controlled by Internet of Things Based on Dynamic Manufacturing Processes’. *Advances in Intelligent Systems and Computing*, 2019, no. 871, pp. 535 – 549, https://doi.org/10.1007/978-3-030-01069-0_38. (індексована в наукометричній базі Scopus, кuartиль Q3 відповідно до класифікації SCImago Journal);
28. Melnykova, Nataliia. ‘Semantic search personalized data as special method of processing medical information’. *Advances in Intelligent Systems and Computing*, 2017, no. 512, pp. 315 – 325, https://doi.org/10.1007/978-3-319-45991-2_22. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
29. Melnykova, Nataliia., et al. ‘The special ways for processing personalized data during voting in elections’. *Advances in Intelligent Systems and Computing*, 1080 AISC, 2020, pp. 781 – 791, DOI: 10.1007/978-3-030-33695-0_52. (індексована в наукометричній базі Scopus);
30. Melnykova, Nataliia., et. al. ‘The personalized approach to the processing and analysis of patients’ medical data’. *CEUR Workshop Proceedings*, 2018, no. 2255, pp. 103 – 112, <http://ceur-ws.org/Vol-2255/paper10.pdf>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
31. Shakhovska, Nataliya. and Melnykova, Nataliia. ‘Feature Engineering and Missing Data Imputation Method of Medical Data Analysis’. *CEUR Workshop Proceedings*, 2022, no. 3137, pp. 48 – 57. <http://ceur-ws.org/Vol-3137/paper4.pdf>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
32. Melnykova, Nataliia., et. al. ‘The problem of analysing the relationships between individual characteristics of individuals with COVID`19’. *CEUR Workshop Proceedings*, 2020, no. 2753, pp. 473 – 482, <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-984587>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
33. Melnykova, Nataliia., et. al. ‘Advisory and accounting tool for safe and economically optimal choice of online self-education services’. *CEUR Workshop Proceedings*, 2019, 2588, pp. 290 – 300, <http://ceur-ws.org/Vol-2588/paper24.pdf>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
34. Melnykova, Nataliia., et. al. ‘Determination of characteristics discrete transfiguration for synthesized raster elements of non-regular structure’. *CEUR Workshop Proceedings*, 2019, no. 2533, pp. 249 – 258, <http://ceur-ws.org/Vol-2533/paper23.pdf>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
35. Melnykova, Nataliia., et. al. ‘Technologies of 3D-prototyping of Objects.’ *CEUR Workshop Proceedings*, 2019, no. 2533, pp. 271 – 281, <http://ceur-ws.org/Vol-2533/paper25.pdf>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);
36. Melnykova, Nataliia. ‘Application of information technology for designing of treatment information systems’. 2015, *Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015*, pp. 156 – 158, <https://doi.org/10.1109/CADSM.2015.7230823>. (індексована в наукометричній базі Scopus та Web of Science Core Collection);

37. Melnykova, Nataliia, and Oksana Markiv. 'Semantic approach to personalization of medical data'. 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT). IEEE, 2016., pp. 59 – 61, <https://doi.org/10.1109/STC-CSIT.2016.7589868>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
38. Melnykova, Nataliia. 'The basic approaches to automation of management by enterprise finances'. 2017, Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017, pp. 288 – 291, <https://doi.org/10.1109/STC-CSIT.2017.8098788>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
39. Melnykova, Nataliia. et. al. 'The personalized approach in a medical decentralized diagnostic and treatment'. 2017, 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2017, pp. 295 – 297, <https://doi.org/10.1109/CADSM.2017.7916139>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
40. Melnykova, Nataliia. et. al. 'The new approaches of heterogeneous data consolidation'. 2018 International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2018, pp. 408 – 411, <https://doi.org/10.1109/STC-CSIT.2018.8526677>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
41. Melnykova, Nataliia. et. al. 'The special ways of application of neural networks for medical information processing, 2018 International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2018, pp. 428 – 431, <https://doi.org/10.1109/STC-CSIT.2018.8526708>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
42. Melnykova, Nataliia. et. al. 'Calculation of the Exact Value of the Fractal Dimension in the Time Series for the Box-Counting Method'. 2019 9th International Conference on Advanced Computer Information Technologies, ACIT 2019, pp. 248 – 251, <https://doi.org/10.1109/ACITT.2019.8780028>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
43. Melnykova, Nataliia. et. al. 'The Applying Processing Intelligence Methods for Classify Persons in Identify Personalized Medication Decisions'. 2020 10th International Conference on Advanced Computer Information Technologies, ACIT 2020, pp. 422 – 425, <https://doi.org/10.1109/ACIT49673.2020.9208822>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
44. Melnykova, Nataliia. et. al. 'The Investigation of Artificial Intelligence Methods for Identifying States and Analyzing System Transitions between States'. 2020 International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2020, vol. 1, pp. 41 – 75 <https://doi.org/10.1109/CSIT49958.2020.9321841>. (індексована в наукометричній базі *Scopus* та *Web of Science Core Collection*);
45. Melnykova, Nataliia. et. al. 'The Scheme Application of the Medical Expert System'. XXIV Ukrainian-Polish Conference on CAD in Machinery Design. Implementation and Educational Issues, CADMD 2016, Lviv, 21.10-22.10.2016, pp. 65 – 66.;
46. Melnykova, Nataliia. et. al. 'The Intelligent System Architecture of Personalized Management'. XXIV Ukrainian-Polish Conference on CAD in Machinery Design,

- Implementation and Educational Issues, CADMD 2016, Lviv, 21.10-22.10.2016, pp.27 – 28.;
47. Мельникова Н.І., Жилко І.В. ‘Застосування хмарних обчислень для проектування систем підтримки лікарських рішень’. Міжнародна конференція «Інноваційні підходи і сучасна наука», Центр наукових публікацій, Київ, 30 квітня 2015 р., pp. 45 – 46.;
 48. Мельникова Н.І., Данилів В.М. ‘Інтелектуальна інформаційна система організації інтерактивних промоакцій’. Матеріалами міжнародної науково-практичної конференції: «IV осінні наукові читання», Київ: збірник статей (рівень стандарту, академічний рівень), Центр наукових публікацій, 2015. pp. 48 – 49.;
 49. Мельникова Н.І., Копач М.І. ‘Методи оптимізації аналізу персоналізованих даних підприємства Математика’. Інформаційні технології. Освіта : тези доповідей VI Міжнародної науково-практичної конференції, Східноєвропейський національний університет ім. Лесі Українки, Луцьк-Світязь, 5 – 7 червня 2017, pp. 67 – 69.;
 50. Мельникова Н.І., Мукалов П., Козій Д. ‘Особливості застосування нейронних мереж щодо опрацювання даних різного походження’. Матеріали VI Всеукраїнської науково-практичної конференції «Наукові дослідження: перспективи інновацій у суспільстві і розвитку технологій», Наукове партнерство «Центр наукових технологій», Харків, 24 – 25 лютого 2017, pp. 104.;
 51. Мельникова Н.І., Мельников В.А. ‘Персоналізований підхід до обробки та аналізу медичних даних пацієнтів’. Матеріали Міжнародної науково-практичної конференції «Інформаційна безпека та інформаційні технології», ХНЕУ імені Семена Кузнеця, 24 – 25 квітня 2019. pp. 56.;
 52. Melnykova, Nataliia. ‘New approach to support of personalized decision making in medicine’. The 8 th International scientific and practical conference « Information, Its Impact on Social And Technical Processes, March Haifa, Israel, 16 – 17, 2020, pp.261.;
 53. Melnykova, Nataliia. et. al. ‘Semantic approach to personalization of medical data’. IX Міжнародна науково-практична конференція «Actual aspects of development in the context of globalization», Florescia, Italy, 23 – 24 March 2020, pp. 126 – 129.;
 54. Мельникова Н.І. ‘Методи оптимізації рішень щодо аналізу персоналізованих даних’. Матеріали II Міжнародної науково-практичної конференції «Інформаційна безпека та інформаційні технології», Кропивницький: ЦНТУ, 2 – 3 квітня 2020, pp. 105.;
 55. Melnykova, Nataliia., and Kolomyi, Anastasia. ‘Information System For Determination Of Early Symptoms Of Dementia Base On Mini-Cog Test And Mini-Mental State Examination Impact of modernity on science and practice’. XVIII International Scientific and Practical Conference, Boston, USA 2020. pp. 95 – 97.;
 56. Melnykova, Nataliia. et. al. ‘Automatic audio to text conversion approaches along the radiology value chain’. XVII International Scientific and Practical Conference «Multidisciplinary academic notes. Theory, methodology and practice», Tokyo, Japan, 03 – 06 May 2022, pp. 994.

АНОТАЦІЇ

Мельникова Н.І. Моделі та методи підтримки персоналізованих рішень у медичних системах. – На правах рукопису.

Дисертаційна робота на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи штучного інтелекту. – Національний університет

«Львівська політехніка», Міністерство освіти і науки України, Львів, 2023.

У дисертаційній роботі вирішено важливу науково-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізація результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

У роботі введено поняття персоналізації, персоналізованих рішень та медичних даних пацієнта. Проаналізовано правові вимоги щодо використання персоналізованих медичних даних, існуючі практичні рішення, що відображають основні підходи до використання медичних даних та урахування персоналізації за результатами аналізу існуючих систем штучного інтелекту у сфері медицини. Проведено порівняльний аналіз класичних методів та визначені обмеження, що характерні при опрацюванні медичних персоналізованих даних. Визначено найбільші обмеження під час опрацювання універсальними методами великих та малих мультимодальних медичних наборів даних, що визначили актуальну науково-прикладну проблему щодо розробки нових чи удосконалення існуючих методів штучного інтелекту щодо підвищення точності процесу обробки великих та малих мультимодальних медичних даних для пошуку персоналізованих рішень та сформуванню завдання наукового дослідження.

У межах дисертаційної роботи введено модель відображення стану пацієнта, яка спрощено відображає його структуру, подає інформацію про його стан та поведінку та представлена як система, що консолідує різні елементи, що представлені у вигляді множин, які взаємозалежні та залежні від середовища оцінки; визначено, що продукційні правила, які формулюють рішення щодо перегляду та зміни тактики лікування; представлено простір стану пацієнта як евклідовий простір, що дозволило змодельовувати інформаційну модель простору станів, як багатовимірну систему в тимчасовій області; формалізовано відображення фізичного стану пацієнта з урахуванням часовозалежних та часовонезалежних даних пацієнта, що дає можливість оцінити його в певний момент часу та приймати сталі показники стану; розроблена модель простору станів пацієнта представлена у вигляді куба, як відображення функціонального відношення загального стану пацієнта; удосконалено метод пошуку зміни стану пацієнта шляхом використання простору умов та аналізу зміни приросту значень часовозалежних даних. Розроблено гібридний ансамбль моделей, що містить кілька селекторів за допомогою агрегування результатів, які будуть використовуватися на етапі попередньої обробки; вбудовані алгоритми виконують вибір ознак під час процедури навчання класифікатора та оптимізують набір ознак, що використовуються для досягнення кращої точності; розроблено метод зменшення розмірності вхідних даних на основі ансамблю моделей слабких предикторів для вибору найважливіших ознак, що базується на прирості інформації з урахуванням результатів застосування декількох селекторів та агрегації кінцевих результатів на підставі урахування індекса Жакара для оцінки подібності одержаних підмножин ознак та проведення мажоритарного голосування результатів. Розроблена трирівнева класифікаційна модель ансамблю стекінгу з агрегатом логістичної регресії, яка має такі значення метрик оцінки продуктивності у вибраній підмножині ознак; проведено порівняння точності прогнозування для стандартних моделей зменшення розмірності під час наступних кроків, для цього використано аналіз головних компонентів із вісьмома компонентами. Розроблено архітектуру системи підтримки прийняття медичних рішень щодо прогнозування станів пацієнта на підставі опрацювання та аналізу персоналізованих медичних даних; подано опис імплементації рішень у системі

підтримки прийняття рішень для лікування хворих з порушеннями антитілоутворень.

Практичне значення результатів дисертаційного дослідження дають змогу підвищити точність прогнозування цільових показників в підмножині простору умов на 5 %, підвищити точність класифікації на 4 % порівняно з результатами логістичної регресії як кращого класифікатора у порівнянні з існуючими для набору даних по COVID - 19 та на 6 %, підвищити точність прогнозування даних на 7-9 % та забезпечити паралелізацію процесу обробки даних як малої, так і великої розмірності, забезпечити заповнення пропусків даних за рахунок одержання додаткових значень а також зменшити імовірність появи похибки (кількість ліжкоднів) при пошуку шаблонів динаміки стану пацієнта шляхом використання простору умов.

Достовірність наукових та практичних результатів підтверджується відповідними матеріалами про впровадження дисертаційних досліджень, а також за рахунок порівняння одержаних практичних результатів з результатами застосування існуючих класичних методів та підходів щодо опрацювання персоналізованої інформації про пацієнта.

Ключові слова: персоналізовані медичні рішення, ієрархічний предиктор, залежності між цільовими змінними, стекінгова модель, заповнення відсутніх даних, пошук зміни стану, модель вибору ознак, великі набори даних, консолідація мультимодальних даних, інформаційна модель стану пацієнта.

Melnykova N.I. Models and methods for supporting personalized solutions in medical systems. – On manuscript rights.

Dissertation for obtaining a technical science doctorate on speciality 05.13.23 – artificial intelligence systems. – Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2023.

The dissertation solved an important scientific and applied problem of developing and improving models, methods and tools of machine learning in the tasks of classification, clustering, forecasting and visualization of the results of personal data processing for adapting medical solutions to the patient.

The work introduces the concepts of personalization, personalized solutions, and patient medical data. Legal requirements for the use of personalized medical data, existing practical solutions reflecting the main approaches to the use of medical data, and taking into account personalization based on the analysis of existing artificial intelligence systems in the field of medicine are analyzed. A comparative analysis of classical methods was carried out and the limitations characteristic of the processing of personalized medical data were determined. The biggest limitations during the processing of large and small multimodal medical data sets by universal methods were determined, which identified an actual scientific and applied problem regarding the development of new or improvement of existing methods of artificial intelligence to increase the accuracy of the process of processing large and small multimodal medical data to find personalized solutions and form tasks scientific research.

Within the scope of the dissertation work, a model for displaying the patient's condition was introduced, which simply reflects its structure, provides information about its condition and behavior, and is presented as a system that consolidates various elements, presented in the form of sets, which are interdependent and dependent on the assessment environment; it is determined that the production rules, which formulate decisions regarding the review and change of treatment tactics; the state space of the patient is presented as a Euclidean space, which made it possible to simulate the information model of the state space as a

multidimensional system in the time domain; the display of the patient's physical condition is formalized, taking into account the patient's time-dependent and time-independent data, which makes it possible to evaluate him at a certain point in time and take constant state indicators; the developed model of the space of the patient's states is presented in the form of a cube, as a reflection of the functional relationship of the patient's general state; the method of searching for a change in the patient's condition by using the space of conditions and analyzing the change in the increment of time-dependent data values has been improved. A hybrid ensemble of models containing several selectors with the help of aggregation of results, which will be used at the stage of pre-processing, has been developed; built-in algorithms perform feature selection during the classifier training procedure and optimize the set of features used to achieve better accuracy; a method of reducing the dimensionality of input data based on an ensemble of weak predictor models for selecting the most important features was developed, which is based on the increase of information taking into account the results of the application of several selectors and the aggregation of the final results based on the consideration of the Jacquard index to assess the similarity of the obtained subsets of features and conduct majority voting of the results; a model of three-layer stacking of an ensemble of methods is developed and the stages of stacking are described, which provides the opportunity to combine associative classification with weak classifiers into an ensemble to generalize the results. A three-level ensemble stacking classification model with a logistic regression aggregate has been developed, which has the following values of performance assessment metrics in the selected subset of features; compared the prediction accuracy of standard dimensionality reduction models in subsequent steps using principal component analysis with eight components. The architecture of the medical decision-making support system for predicting the patient's condition based on the processing and analysis of personalized medical data has been developed; a description of the implementation of decisions in the decision-making support system for the treatment of patients with disorders of antibody formation is presented.

The practical value of the results of the dissertation research makes it possible to increase the accuracy of forecasting targets in a subset of the conditioned space by 5%, to increase the classification accuracy by 4% compared to the results of logistic regression as a better classifier compared to the existing ones for the COVID - 19 data set and by 6%, increase the accuracy of data forecasting by 7-9% and ensure the parallelization of the data processing process of both small and large dimensions, ensure the filling of data gaps due to obtaining additional values, and also reduce the probability of the appearance of an error (number of bed days) when searching for patterns of the dynamics of the patient's condition by using the space of conditions.

Relevant materials confirm the validity of scientific and practical results on the implementation of dissertation studies, as well as by comparing the obtained empirical results with the results of the application of existing classical methods and approaches for processing personalized information about the patient.

Keywords: personalized medical decisions, hierarchical predictor, dependencies between target variables, stacking model, missing data filling, state change search, feature selection model, large data sets, multimodal data consolidation, patient state information model.