

Міністерство освіти і науки України  
Національний університет “Львівська політехніка”

ІЗОНІН ІВАН ВІКТОРОВИЧ



УДК 004.85:004.32.26:004.032.2](043)

**АНСАМБЛІ МОДЕЛЕЙ НЕІТЕРАТИВНОГО МАШИННОГО  
НАВЧАННЯ ДЛЯ АНАЛІЗУ БІОМЕДИЧНИХ ДАНИХ  
МАЛИХ ОБСЯГІВ**

05.13.23 – Системи та засоби штучного інтелекту  
технічні науки

Реферат дисертації на здобуття наукового ступеня доктора технічних наук

Львів – 2025

Дисертацією є рукопис.

Робота виконана в Національному університеті “Львівська політехніка” Міністерства освіти і науки України.

**Науковий консультант:** доктор технічних наук, професор  
**Ткаченко Роман Олексійович**,  
професор кафедри інформаційних технологій  
видавничої справи  
Національного університету  
“Львівська політехніка”.

**Офіційні опоненти:** доктор технічних наук, професор  
**Стіренко Сергій Григорович**,  
проректор з наукової роботи  
Національного технічного університету України  
“Київський політехнічний інститут ім. Ігоря  
Сікорського”;

доктор технічних наук, професор  
**Антощук Світлана Григорівна**,  
директор Інституту комп’ютерних систем  
Національного університету  
“Одеська політехніка”;

доктор технічних наук, професор  
**Субботін Сергій Олександрович**,  
завідувач кафедри програмних засобів  
Національного університету  
“Запорізька політехніка”.

Захист відбудеться “26” серпня 2025 р. о 12.00 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 та на сайті докторської ради Д 35.052.14 Національного університету “Львівська політехніка” за покликанням: <https://lpnu.ua/spetsrady/d-3505214>

Учений секретар докторської ради Д 35.052.14,  
кандидат технічних наук, доцент

Віктор ХАВАЛКО

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Актуальність теми дослідження зумовлена зростаючою потребою в ефективних та стійких методах аналізу біомедичних даних, які в реальних умовах часто представлені у вигляді вибірок даних малого обсягу. Останні характеризуються гетерогенністю, наявністю шумів, пропусків і аномалій, а також значною вимірністю. Класичні статистичні методи та більшість сучасних моделей машинного навчання, включаючи штучні нейронні мережі, не забезпечують достатнього рівня властивості узагальнення на нових даних в таких умовах функціонування. Це значно обмежує їх використання у практиці біомедичних досліджень та клінічної діагностики. Особливої важливості ця проблема набуває в задачах побудови моделей регресії чи класифікації на основі біомедичних вибірок даних малих обсягів, що характерно для рідкісних захворювань, експериментальних методів лікування або персоналізованих протоколів терапії.

Проблема аналізу малих біомедичних вибірок особливо актуальна у зв'язку з високою вартістю, часозатратністю, етичними обмеженнями, рідкісністю досліджуваних станів або складністю експериментального відтворення клінічних даних. Існуючі методи нелінійного збільшення вимірності простору ознак, аугментації даних, синтезу ансамблів штучних нейронних мереж та компенсації похибок ансамблів моделей не завжди забезпечують достатню точність, стійкість до варіативності даних, а інколи – і можливість їх практичного використання в умовах аналізу біомедичних даних малого обсягу. Це зумовлює необхідність удосконалення існуючих та розроблення нових методів аналізу таких даних. Значної уваги потребують неітеративні моделі, які дозволяють зменшити обчислювальну складність без втрати точності та забезпечити прозорість і повторюваність результатів. У цьому напрямку особливу цінність мають методи, які забезпечать можливість розв'язання задач регресії та класифікації в умовах обмеженого обсягу даних для реалізації методів навчання, зокрема, штучних нейронних мереж.

Проблемами інтелектуального аналізу даних, зокрема, з використанням нейромережових засобів активно займалися такі українські вчені: С. Г. Антощук, С. А. Бабічев, О. В. Бармак, О. М. Березький, О. В. Бісікало, Є. В. Бодянський, О. А. Винокурова, Т. О. Говорущенко, В. А. Гуляєв, Ю. П. Зайченко, О. Г. Івахненко, Р. М. Камінський, В. В. Ковтун, Ю. В. Крак, В. І. Литвиненко, Н. І. Мельникова, А. О. Олійник, І. Г. Перова, С. Г. Стіренко, С. О. Субботін, Р. О. Ткаченко, І. Г. Цмоць, А. Ю. Шафроненко, Н. Б. Шаховська, В. С. Яковина, а також ряд закордонних вчених: D. Estrin, E. B. Hekler, J. Jiang, R. Kitchin, P. Kokol, P. Klasnja, C. Li, T. Shaikhina, P. Xu. Зокрема, О. Г. Івахненко здійснив фундаментальні дослідження у галузі синтезу нейронних мереж і став автором методу групового урахування аргументів, який поклав початок автоматичному моделюванню складних систем; методами та засобами опрацювання діагностичної інформації в реальному часі займався В. А. Гуляєв;

неітеративними методами синтезу нейромережових моделей – Р. О. Ткаченко, С. О. Субботін; проблемами апаратної реалізації таких нейронних мереж – І. Г. Цмоць; теоретичним обґрунтуванням необхідності розвитку парадигми малих даних – Е. В. Неклер та Р. Класнґа; проблемами розпізнавання образів та кластеризації на даних малих обсягів – Є. В. Бодянський, Р. М. Камінський, О. А. Винокурова, І. Г. Перова, А. Ю. Шафроненко; алгебраїчним підходом до аналізу даних малих обсягів – Ю. В. Крак; задачами оцінювання параметрів стохастичних моделей на основі вибірок даних малих обсягів – О. В. Бісікало, В. В. Ковтун; проблемами аналізу малих даних засобами глибокого навчання та пояснювального штучного інтелекту – О. В. Бармак, Ю. В. Крак; проблемами аналізу високовимірних даних у галузі біоінформатики – В. І. Литвиненко, С. А. Бабічев; проблемами персоналізованої медицини – Ю. П. Зайченко, Н. І. Мельникова; проблемами аугментації малих обсягів даних та їх класифікації – О. М. Березький, С. Г. Стіренко; проблемами ансамблевого навчання – Н. Б. Шаховська, С. Г. Стіренко, С. Г. Антощук, В. С. Яковина.

У задачах аналізу структурованих біомедичних даних на вибірках малих обсягів використання окремих моделей машинного навчання, зокрема, класичних штучних нейронних мереж (ШНМ) обмежене через високу ймовірність перенавчання, недостатню репрезентативність вибірок для коректного налаштування параметрів моделі та низьку здатність до узагальнення. Особливо це стосується нелінійних моделей, які мають велику кількість параметрів та потребують значного обсягу інформації для ефективної реалізації процедур навчання. Використання ансамблів ШНМ, може частково вирішити проблему перенавчання завдяки зменшенню дисперсії оцінок і покращенню властивостей узагальнення моделі. Однак синтез ансамблів гетерогенних ШНМ, який вимагає наявності різних базових моделей з хоча б прийнятною точністю, є дуже складним, а іноді – практично недосяжним, зокрема, у випадку аналізу екстремально-коротких вибірок даних. Окрім цього, ітеративні методи навчання існуючих ШНМ є ресурсомісткими, повільними та залежними від початкової рандомізації. Це знижує відтворюваність отриманих рішень і ускладнює їх використання в умовах обмежених обчислювальних можливостей. Сукупно перелічені фактори обумовлюють необхідність удосконалення існуючих та розроблення нових, точніших, стійких до перенавчання і обчислювально ефективних ансамблів моделей неітеративного машинного навчання для аналізу структурованих біомедичних вибірок даних малих обсягів.

**Зв'язок роботи з науковими програмами та темами.** Дисертаційна робота виконана на кафедрі систем штучного інтелекту Національного університету “Львівська політехніка” та відповідає науковому напрямку кафедри “Методи та засоби штучного інтелекту для підтримки прийняття рішень”. Основні теоретичні та прикладні результати дисертації отримано в процесі участі (як керівник чи виконавець) у виконанні ряду науково-дослідних робіт як національного, так і міжнародного рівня, а саме: наукового

гранту EURIZON: “Artificial Intelligence in Medicine: Small Data Approach (SmallAIM)” (№ 871072, 01.04.2024 – 31.03.2025, фінансування: Європейський Союз, програма Horizon Europe); прикладного дослідження “Методи та засоби дослідження маркерів старіння та їх впливу на постковідні ефекти для продовження працездатного періоду” (№ 2021.01/0103, 2022 – 2024, фінансування: Національний фонд досліджень України); науково-технічної (експериментальної) розробки “Методи та засоби інформаційної безпеки та гігієни на основі інтерпретованого штучного інтелекту” (№ 0121U107809, 2022 – 2024, фінансування: Міністерство освіти і науки України); науково-технічної (експериментальної) розробки “Технології та системи оброблення і зберігання персоналізованих військових медичних даних” (№ 0121U107809, 2022 – 2023, фінансування: Міністерство освіти і науки України); науково-технічної (експериментальної) розробки “Розроблення інформаційної технології оцінювання та прогнозування надійності програмного забезпечення методами машинного навчання” (№ 0121U109527, 2021 – 2022, фінансування: Міністерство освіти і науки України); наукового гранту “Integrated care for next generation” (iCare4NextG) (C2020/1-8, 04.2021 – 03.2024, фінансування: Європейський Союз, програма EUREKA); прикладного дослідження “Система підтримки прийняття рішень моделювання поширення вірусних інфекцій” (№ 211/01.220, 2020 – 2021, фінансування: Національний фонд досліджень України); прикладного дослідження “Національний реєстр пацієнтів на первинні імунodefіцити” (№ 0121U107803, 2021 – 2023, фінансування: Міністерство охорони здоров’я України); прикладного дослідження “Інформаційна технологія опрацювання персоналізованої медичної інформації” (№ 0119U002257, 2019 – 2020, фінансування: Міністерство освіти і науки України).

**Мета дослідження** – розвиток основ теорії синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для підвищення точності розв’язування задач регресії та класифікації на вибірках біомедичних даних малих обсягів.

**Завдання дослідження:**

1. Проаналізувати методи, моделі та засоби інтелектуального опрацювання біомедичних даних малих обсягів.
2. Розвинути метод кусково-лінійної апроксимації для моделювання багатовимірних поверхонь відгуку.
3. Удосконалити метод формування комбінованої поверхні відгуку меншої вимірності з застосуванням кластерного аналізу на вибірках даних малих обсягів.
4. Розробити метод нелінійного доповнення простору незалежних ознак біомедичної вибірки даних вихідними сигналами шару підсумовування ймовірнісної нейронної мережі.
5. Розробити методологію розв’язання задач регресії та класифікації на екстремально-коротких біомедичних вибірках з використанням аугментації даних.

6. Розробити метод синтезу ансамблів лінійних нейроподібних структур з неітеративним машинним навчанням з використанням різних підходів до агрегування кусково-лінійно апроксимованих поверхонь відгуку кожного його елемента.

7. Розробити метод синтезу ансамблю гомогенних нелінійних ШНМ, що функціонують на незалежно сформованих вибірках даних, з використанням симетричних випадкових зміщень та з різними способами агрегування вихідних результатів окремих моделей.

8. Розробити метод часткової компенсації систематичної похибки нейронної мережі узагальненої регресії.

9. Удосконалити метод глобально-локальної апроксимації шляхом незалежного виділення глобальної та локальної компонент сигналу ансамблю нейронних мереж.

10. Розробити прямий та зворотний методи компенсації похибок регресійних моделей з використанням каскадних ансамблів на основі раціональних дробів.

11. Розвинути теорію синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням на вибірках біомедичних даних малих обсягів.

12. Розробити програмне забезпечення для реалізації усіх методів та виконати оцінювання їх результатів на основі як модельних даних, так і даних біомедичних досліджень.

**Об'єкт дослідження** – процеси розроблення методів синтезу нейроподібних ансамблевих структур для інтелектуального аналізу структурованих вибірок даних малого обсягу.

**Предмет дослідження** – моделі, методи та засоби синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для розв'язання задач регресії та класифікації на біомедичних вибірках даних малого обсягу.

**Методи дослідження.** У дисертаційній роботі використано методи математичного аналізу та чисельного моделювання для формалізації, побудови та теоретичного обґрунтування нових ансамблів ШНМ, методи теорії ймовірностей і математичної статистики для оцінювання точності та здатності до узагальнення розроблених методів, а також для визначення критеріїв зупинки і перевірки їх ефективності в умовах обмежених вибірок даних, методи теорії ШНМ і ансамблевого навчання для побудови неітеративних ансамблів ШНМ, розроблення методів їх навчання, а також реалізації механізмів компенсації похибок, методи структурного й компонентного аналізу даних, які використовувалися для збільшення вимірності простору ознак з залученням полінома Колмогорова-Габора, декартового квадрата та ізоморфного представлення вхідних компонент векторів, методи кластерного аналізу для побудови комбінованих поверхонь відгуку та реалізації аугментації даних на основі кластеризації. Методи комп'ютерного моделювання та експериментальної перевірки

використовувалися для реалізації удосконалених та розроблених методів, проведення експериментів і виконання порівняльного аналізу ефективності моделей у задачах регресії та класифікації на біомедичних вибірках даних малих обсягів.

**Наукова новизна роботи.** У дисертаційній роботі на основі виконання теоретичних та експериментальних досліджень вирішено важливу теоретичну проблему розвитку основ теорії синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для підвищення точності розв'язання задач регресії та класифікації на вибірках біомедичних даних малих обсягів. При цьому отримано такі нові наукові результати:

*уперше розроблено:*

- методологію розв'язання задач регресії й бінарної класифікації на основі декартового квадрата множини вхідних компонент векторів, яка містить базовий, адитивний, каскадний методи, а також методи формування комбінованої поверхні відгуку та формування аугментованої вибірки даних необхідного обсягу, що забезпечує підвищення як точності роботи нелінійних моделей машинного навчання, так і їх властивості узагальнення в умовах навчання на екстремально-коротких вибірках біомедичних даних;
- метод синтезу ансамблю лінійних нейроподібних структур, який завдяки агрегуванню кусково-лінійно апроксимованих поверхонь відгуку кожного його елемента забезпечує підвищення точності моделей регресії та класифікації;
- метод синтезу ансамблів гомогенних нелінійних штучних нейронних мереж, який шляхом агрегування апроксимованих поверхонь відгуку, сформованих на незалежних вибірках із використанням попарних симетричних випадкових зміщень, підвищує точність моделей регресії та класифікації;
- метод часткової компенсації систематичної похибки нейронної мережі узагальненої регресії, який за рахунок використання ансамблю з двох таких нейронних мереж забезпечує підвищення точності розв'язання задач регресії у випадку аналізу біомедичних даних малих обсягів;
- метод нелінійного розширення простору незалежних ознак біомедичної вибірки даних, який внаслідок доповнення їх вихідними сигналами шару підсумовування ймовірнісної нейронної мережі забезпечує підвищення точності класифікації даних на лінійних неітеративних нейроподібних структурах;
- прямий та зворотний методи компенсації похибки моделі регресії, у яких за рахунок перетворень на основі раціональних дробів досягнуто підвищення точності та робастності апроксимації в задачах аналізу біомедичних даних малих обсягів з можливістю додаткового ітеративного уточнення результатів;

*вдосконалено:*

- метод глобально-локальної апроксимації біомедичних даних малих обсягів, який на відміну від наявного використовує незалежне виділення глобальної та локальної компонент сигналу з використанням ансамблю з двох нейронних мереж узагальненої регресії та надає змогу додатково ітеративно уточнити шукане значення;

- метод формування комбінованої поверхні відгуку з застосуванням кластерного аналізу, який на відміну від наявного доповнює простір ознак значенням залежного атрибута центру кластера, що забезпечує підвищення точності моделей регресії та класифікації на вибірках даних малих обсягів без втрати узагальнення;

*набули подальшого розвитку:*

- теорія синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням, яка ґрунтується на розроблених методах нелінійного збільшення вимірності простору ознак, аугментації даних, декореляції вибірок та компенсації похибок для підвищення точності розв'язання задач регресії та класифікації на вибірках біомедичних даних малих обсягів;

- метод кусково-лінійної апроксимації, який за рахунок ізоморфного подання кожного значення ознаки у вигляді її локальних компонент збільшує вимірність простору вхідних даних та забезпечує локальну корекцію параметрів багатовимірної поверхні відгуку лінійними неітеративними нейроподібними структурами.

**Практичне значення** одержаних результатів полягає у тому, що розроблені та вдосконалені методи і програмні засоби є ефективними інструментами для інтелектуального аналізу біомедичних даних малого обсягу. Ці методи відзначаються високими показниками точності, робастності до шумів та аномалій, а також обчислювальною ефективністю. Використання розроблених та удосконалених методів забезпечує:

- суттєве підвищення точності лінійних моделей машинного навчання, зберігаючи їх здатність до узагальнення за рахунок незначного збільшення вимірності простору незалежних ознак кожного вектора даних;

- підвищення точності нелінійних моделей машинного навчання до 59% та їх стійкості до спотворень у даних шляхом використання розробленої методології розв'язання задач регресії та бінарної класифікації на основі декартового квадрата множини вхідних компонент векторів;

- збільшення точності на 5 – 6% в умовах аналізу даних малих обсягів на основі розроблених ансамблів лінійних і нелінійних ШНМ з застосуванням різних способів декореляції вхідних даних та агрегування результатів;

- зменшення середньоквадратичної похибки на 4 – 8% залежно від способу агрегування результатів роботи методу часткового усунення систематичної похибки нейронної мережі узагальненої регресії;

- зменшення середньоквадратичної похибки результатів розв'язання задачі регресії на 13% з використанням удосконаленого методу глобально-

локальної апроксимації та на 9% і 11% відповідно з використанням прямого і зворотного методів компенсації похибок на основі раціональних дробів.

Розроблені програмні засоби, захищені авторським правом на комп'ютерні програми, забезпечили можливість експериментального оцінювання ефективності усіх методів як на модельних, так і на реальних даних біомедичних досліджень. Отримані результати успішно апробовані при розв'язанні низки прикладних задач у різних напрямках біомедичних досліджень, зокрема, в біоматеріалознавстві, ревматології, травматології, стоматології та трансплантології. Їх впровадження здійснювалося у співпраці з медичними установами та науково-дослідними організаціями, а також в межах виконання наукового-дослідних робіт в Україні та міжнародних грантових проєктів за кордоном, що засвідчено відповідними актами впровадження та використання. Зокрема, результати проведених досліджень впроваджено: у Військово-медичному клінічному центрі Західного регіону в період з січня 2024 по січень 2025 року; у Комунальному некомерційному підприємстві Харківської обласної ради “Обласний медичний клінічний центр урології і нефрології ім. В. І. Шаповала” у 2022 – 2024 роках; на кафедрі стоматології дитячого віку та імплантології Харківського національного медичного університету у 2022 – 2024 роках; у Комунальному некомерційному підприємстві Харківської обласної ради “Обласна клінічна лікарня” у 2022 – 2024 роках; у межах виконання міжнародного наукового гранту EURIZON “Artificial Intelligence in Medicine: Small Data Approach (SmallAIM)” (№ 871072, 2024 – 2025); у межах виконання науково-дослідних робіт в Україні, які фінансувалися Міністерством освіти і науки України, а також Національним Фондом Досліджень України.

Окрім цього, розроблені та удосконалені методи інтелектуального аналізу біомедичних даних малих обсягів можуть використовуватися для розв'язання задач, що виходять за межі опрацювання суто біомедичних даних малих обсягів. Завдяки універсальності закладених підходів, розроблені та удосконалені у дисертації методи можуть бути успішно адаптовані та ефективно використані для інтелектуального аналізу даних у суміжних прикладних галузях, де також спостерігаються вибірки даних малих обсягів.

**Особистий внесок здобувача.** Усі наукові результати, подані в дисертації, є особистим досягненням здобувача. Усі положення, що виносяться на захист, належать авторові особисто. Наукові праці [1], [3] виконані одноосібно. У працях, опублікованих у співавторстві, концептуальні ідеї, теоретичні підходи та методи, які покладено в основу дисертаційного дослідження, розроблені здобувачем самостійно. Відповідно до пунктів наукової новизни дисертаційної роботи здобувачем зроблено такі внески: розроблено методологію розв'язання задач регресії і бінарної класифікації на основі декартового квадрата множини вхідних компонент векторів [23], [26], яка включає розроблення базового [36], [37], [38], [40], адитивного [11], [12], каскадного [13], [15] методів, а також розроблення методів формування комбінованої поверхні відгуку [14] та аугментованої вибірки даних

зменшеного обсягу [41]; розвинуто метод кусково-лінійної апроксимації для задач регресії і класифікації [33], [35], [47] та розроблено метод синтезу ансамблю лінійних нейроподібних структур на його основі [25], [43]; розроблено методи синтезу ансамблів гомогенних нелінійних ШНМ та нейроподібних структур з розширенням входів радіально-базисними функціями (РБФ) з різними способами формування підвбірок даних та агрегування результатів [2], [4], [5], [6], [42], [44], [45], [46]; розроблено метод часткового усунення систематичної похибки нейронної мережі узагальненої регресії з використанням трьох різних способів агрегування результатів [22], [24], [50]; розроблено метод нелінійного доповнення простору незалежних ознак біомедичної вибірки даних вихідними сигналами шару підсумовування різними модифікаціями ймовірнісної нейронної мережі [8], [10], [28]; розроблено прямий та зворотний методи компенсації похибки моделі регресії з можливістю ітеративного уточнення результатів [48], [49]; удосконалено метод глобально-локальної апроксимації біомедичних даних малих обсягів [16], [27]; удосконалено метод формування комбінованої поверхні відгуку з застосуванням кластерного аналізу [30], [32]; розвинуто теорію синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням, а саме: розроблено метод нелінійного доповнення входів ШНМ поліномом Колмогорова-Габора [7], розроблено двокроковий метод нормалізації табличних вибірок даних [9], розроблено каскадний ансамбль неітеративних ШНМ на основі використання полінома Колмогорова-Габора та метод його поліноміального подання у режимі застосування [17], розроблено метод зменшення вимірності простору вхідних даних для кожного члена каскадного ансамблю [18], розроблено метод зменшення тривалості процедури навчання каскадного ансамблю [19], розроблено метод каскадування алгоритмів машинного навчання [20], досліджено особливості функціонування ймовірнісної нейронної мережі в процесі аналізу даних малих обсягів [29], досліджено вплив ефективності використання різних методів нормалізації даних на точність інтелектуального аналізу даних малих обсягів [31], розроблено метод розширення входів нейронної мережі узагальненої регресії [34], досліджено вплив ефективності існуючих методів аугментації табличних вибірок даних на точність роботи алгоритмів машинного навчання [39], розроблено каскадний ансамбль неітеративних штучних нейронних мереж на основі принципу лінеаризації поверхні відгуку [50], розроблено програмні засоби для реалізації усіх методів [51], [52], [53].

**Апробація результатів дисертації.** Основні теоретичні та прикладні результати дисертаційної роботи широко обговорені на провідних науково-практичних конференціях, симпозіумах і семінарах, що відбувалися в 13 країнах світу, а саме: XVI Міжнародній науково-практичній конференції “Інформаційні технології і автоматизація – 2023” (жовтень 2023, Одеса, Україна), 15-тій Ювілейній Міжнародній науково-практичній конференції молодих вчених “Економічний і соціальний розвиток України в XXI столітті: національна візія та виклики глобалізації” (травень 2023, Тернопіль, Україна),

XXI-й Міжнародній науковій конференції “Нейромережні технології та їх застосування” (HMT3–2023) (2023, Краматорськ–Вінниця–Тернопіль, Україна), 33th International Conference on Database and Expert Systems Applications (DEXA 2022) (22–24 серпня 2022, Відень, Австрія), International Conference on Informatics & Data-Driven Medicine (IDDM 2020) (19–21 листопада 2020, Векше, Швеція), IEEE International Conference on Electronics and Information Technologies (ELIT 2023) (26–28 вересня 2023, Львів, Україна), International Conference on Advanced Computer Information Technologies (ACIT 2022) (26–28 вересня 2022, Списька Капітула, Словаччина), International Conference on Decision Aid Sciences and Application (DASA 2020) (8–9 листопада 2020, Сахір, Королівство Бахрейн), International Work-Conference on Artificial Neural Networks (IWANN 2019) (12–14 червня 2019, Гран-Канарія, Іспанія), 6th International Symposium on Emerging Information, Communication and Networks (EICN 2019) (4–7 листопада 2019, Коїмбра, Португалія), 2-nd International Conference on Computer Science, Engineering and Education Applications (ICCSEEA2019) (26–27 січня 2019, Київ, Україна), International Conference on Artificial Intelligence and Logistics Engineering (ICAILE2021) (22–24 січня 2021, Київ, Україна), 12th International Conference on Ambient Systems, Networks and Technologies (ANT 2021) (23–26 березня 2021, Варшава, Польща), 4th International Conference on Emerging Data and Industry 4.0 (EDI40 2021) (23–26 березня 2021, Варшава, Польща), International Conference on Informatics & Data-Driven Medicine (IDDM 2021) (19–21 листопада 2021, Валенсія, Іспанія), International Conference on Data Science and Applications (ICDSA 2021) (10–11 квітня 2021, Калькутта, Індія), International Conference on Future Networks and Communications / Mobile Systems and Pervasive Computing (FNC 2024) (5–7 серпня 2024, Гантінгтон, Західна Вірджинія, США), 2-nd International Workshop on Reliability Engineering and Computational Intelligence (RECI 2022) (14–15 листопада 2022, Жиліна, Словаччина), International Conference on Big Data Innovations and Applications (Innovate-Data 2019) (2019, Стамбул, Туреччина), International Conference on Mobile Web and Intelligent Information Systems (MobiWIS 2019) (2019, Стамбул, Туреччина), IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON 2019) (4–6 липня 2019, Львів, Україна), International Workshop on Informatics & Data-Driven Medicine (IDDM 2018) (листопад 2018, Львів, Україна), IEEE International Conference on Computer Science and Information Technologies (CSIT 2023) (18–20 жовтня 2023, Львів, Україна), Mediterranean Smart Cities Conference (MSCC 2024) (травень 2024, Мартіль–Тетуан, Марокко) та International Workshop on Small and Big Data Approaches in Healthcare (SBDaH 2021) (1–4 листопада 2021, Льовен, Бельгія).

**Публікації.** Основні результати дисертаційного дослідження опубліковано у 53 наукових публікаціях. Зокрема, опубліковано два розділи у колективних монографіях в Україні та за кордоном, 22 статті у наукових журналах, з яких 6 – у фахових виданнях України категорії Б, а 16 – у високореєтингових закордонних журналах, що індексуються у

наукометричних базах Scopus та/або Web of Science. Серед них чотири статті опубліковано у журналах першого квартилю (Q1), а одинадцять – у журналах другого квартилю (Q2). Крім того, матеріали дисертації висвітлено у 26 публікаціях у збірниках матеріалів міжнародних наукових конференцій, з яких 22 індексуються у наукометричній базі Scopus. За результатами дослідження отримано три свідоцтва України про реєстрацію авторського права на комп'ютерні програми.

**Структура та обсяг роботи.** Дисертаційна робота складається з вступу, семи розділів, висновків, переліку використаних джерел, що включає 310 найменувань, та п'яти додатків. Загальний обсяг дисертації становить 371 сторінку, з яких основний текст займає 310 сторінок. Робота містить 50 рисунків і 17 таблиць, що ілюструють основні результати дослідження.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теоретичної проблеми, розв'язанню якої присвячено дослідження, здійснено аналіз сучасного стану розробок у галузі інтелектуального аналізу біомедичних вибірок даних малих обсягів. Сформульовано мету дослідження, визначено об'єкт, предмет та методи дослідження. Визначено ряд завдань, послідовне вирішення яких забезпечує досягнення поставленої мети. Подано зв'язок дисертаційної роботи з виконанням державних і міжнародних науково-дослідних проєктів, в межах яких одержано результати дослідження. Викладено наукову новизну та практичну цінність одержаних результатів, наведено відомості про апробацію основних положень роботи та перелік наукових публікацій за її темою, вказано особистий внесок здобувача.

У **першому розділі** дисертаційної роботи здійснено аналіз сутності та характеристик біомедичних даних, що уможливило виявлення критичних обмежень, властивих задачам за умов опрацювання вибірок даних малих обсягів. Подано визначення малих даних як вибірок даних малого обсягу та вузької спрямованості, які призначені для розв'язання конкретних задач за умов ресурсних, часових, етичних або конфіденційних обмежень і характеризується високим співвідношенням числа ознак до кількості спостережень. Особливої уваги потребують екстремально-короткі вибірки даних (менше 100 спостережень), які, зазвичай, унеможливлюють гарантії узагальнення з використанням існуючих нелінійних методів машинного навчання.

Обґрунтовано доцільність використання лінійних неітеративних нейроподібних структур на основі Моделі Послідовних Геометричних Перетворень (МППГ), які демонструють повторюваність результатів навчання, високу інтерпретованість та швидкодію за рахунок неітеративності алгоритму навчання. Проте для підвищення точності розв'язання задач регресії та класифікації вони потребують нелінійного розширення вхідного простору даних.

Розглянуто властивості нейронної мережі узагальненої регресії (НМУР) та ймовірнісної нейронної мережі (ЙНМ), які показують високу адаптивність у випадках аналізу даних малих обсягів. Однак ефективність цих ШНМ різко знижується у випадку аналізу екстремально-коротких вибірок даних. Проаналізовано потенціал методів аугментації табличних даних, які дозволяють підвищити щільність покриття простору ознак і знизити ризик перенавчання за умови збереження статистичної, структурної та клінічної валідності. Обґрунтовано доцільність застосування різних класів ансамблів моделей ШНМ, які здатні підвищити точність, робастність і властивості узагальнення на нових даних за рахунок поєднання результатів окремих моделей з урахуванням їх різноманіття. Встановлено, що ефективність ансамблів ШНМ в задачах аналізу біомедичних вибірок даних малого обсягу визначається типом базових моделей, способами декореляції навчальних підмножин, механізмами агрегування вихідних сигналів та конфігурацією ансамблю. Існуючі типи ансамблів не завжди демонструють високу ефективність при аналізі даних малих обсягів. Це зумовлює необхідність розробки нових архітектур на основі неітеративного машинного навчання, здатних адекватно враховувати особливості та обмеження відповідного класу задач.

Результати першого розділу висвітлено у наукових публікаціях [9, 11, 17, 20, 26, 31-33, 35, 39, 43].

**Другий розділ** присвячено розробці та дослідженню методів нелінійних перетворень вхідних компонент векторів для формування простору ознак підвищеної вимірності. Формально постановка задачі має такий вигляд. Нехай маємо навчальну вибірку даних малого обсягу:

$$A \in \mathbb{R}^{n \times m}, \quad (1)$$

де  $n$  – кількість векторів,  $m$  – вимірність кожного вектора. При цьому кожен рядок  $A$  позначимо як  $x_i \in \mathbb{R}^m$  для  $i = \overline{1, n}$ , а відповідні значення залежного атрибута  $y_i$  для кожного  $x_i \rightarrow y_i$  подані одним з варіантів:

$$\begin{cases} y_i \in \{1, 2, \dots, K\}, & \text{для задач класифікації;} \\ y_i \in \mathbb{R}, & \text{для задач регресійного аналізу.} \end{cases} \quad (2)$$

Задача полягає у тому, щоб знайти таке нелінійне відображення (функцію ознак), яке перетворюватиме кожен вектор  $x_i$  у новий вектор в просторі вищої вимірності так, щоб залежність між  $x_i$  та  $y_i$  у цьому просторі була максимально лінійною. Проекція  $x_i$  у простір вищої вимірності, зокрема, через подібне відображення значно збільшить ймовірність лінійного розділення даних, які в початковому просторі демонструють складні нелінійні залежності (теорема Ковера).

У дисертації розроблено метод розширення простору ознак за рахунок доповнення кожного вхідного вектора членами полінома Колмогорова-Габора (розклад Іто або поліном Вінера). Ці члени включають степені вхідних змінних та їх добутки. Це дозволяє описувати складні нелінійні залежності, зокрема, у

даних біомедичних досліджень. Перевагою такого підходу є висока апроксимаційна здатність цього полінома як дискретного аналогу ряду Вольтерри. Вона істотно підвищує точність розв'язання задач регресії та класифікації з використанням лінійних методів машинного навчання. У разі застосування лінійної нейроподібної структури МППП на розширеній таким чином вибірці даних, яка забезпечує неітеративний, відтворюваний процес навчання, можна синтезувати коефіцієнти цього полінома за допомогою методу діагональної матриці. Розроблений метод особливо ефективний для задач, близьких до вироджених, де класичний метод найменших квадратів не дає задовільних оцінок параметрів. Саме тому він обмежується вузьким класом задач в умовах аналізу біомедичних даних малих обсягів.

У випадках аналізу складних, високовимірних, неоднорідних біомедичних даних глобальна модель ШНМ в процесі аналізу досліджуваної вибірки даних може не враховувати усіх можливих варіантів залежностей між даними. Це впливатиме на точність її роботи, недостатній рівень якої нівелюватиме можливість використання такої моделі апроксимації чи класифікації даних на практиці. У роботі розроблено метод розв'язання задач класифікації, який ґрунтується на розширенні входів лінійних нейроподібних структур МППП вихідними сигналами шару підсумовування ЙНМ. У цьому шарі ЙНМ формуються ймовірності приналежності до кожного з класів задачі, причому їх можна обчислювати за різними схемами: класичним способом, на основі побудови повної системи подій або з урахуванням незбалансованості вибірки даних.

На основі врахування цих ймовірнісних оцінок розроблено метод і його варіації для збільшення вимірності простору вхідних ознак. Це дозволяє підвищити точність класифікації з використанням лінійних ШНМ. Основний підхід полягає у доповненні кожного вектора ознак початкової вибірки даних набором отриманих ймовірностей належності поточного спостереження до усіх класів задачі з подальшим застосуванням лінійної ШНМ на такій розширеній вибірці. Це дозволяє зберегти інформативність первинних атрибутів і водночас посилити модель додатковими ознаками, які несуть інформацію про ймовірнісну структуру класів. Такий підхід забезпечує врахування взаємозв'язків між вхідними параметрами та ймовірнісними оцінками, підвищуючи лінійну роздільність даних і, відповідно, точність класифікації. Завдяки контрольованому збільшенню вимірності вхідного простору даних метод зберігає обчислювальну ефективність, не виходячи за межі параметричного навантаження, прийнятного для лінійних ШНМ. Це, своєю чергою, забезпечує виконання умов теореми Вапніка–Червоненкіса.

Задля врахування інформації про геометричну структуру даних у дисертаційній роботі удосконалено метод формування комбінованої поверхні відгуку з застосуванням кластерного аналізу для підвищення точності розв'язання задач регресії та класифікації у випадку аналізу даних малих обсягів. Удосконалений метод, на відміну від існуючого, полягає у додаванні лише значення залежного атрибута центра кластера як додаткової ознаки до

відповідного вектора вхідних даних, який належить зазначеному кластеру. Такий підхід дозволяє досягти двох ключових переваг в процесі аналізу високовимірних даних малих обсягів. По-перше, він значно зменшує кількість незалежних змінних, необхідних для подальшого аналізу, що спрощує модель і робить її більш ефективною. По-друге, інтеграція інформації, отриманої через кластеризацію, забезпечує додаткову інформативність для регресора чи класифікатора за рахунок формування таким чином комбінованої поверхні відгуку. Обидві переваги забезпечують підвищення точності роботи лінійних методів і їх здатність до узагальнення, що є важливим для застосування на нових, не бачених раніше даних.

Автором також розвинуто метод кусково-лінійної апроксимації для аналізу багатовимірних поверхонь відгуку. Він збільшує вимірність простору вхідних даних на основі ізоморфного подання кожного одиничного значення ознаки як сукупності локальних ознак і дозволяє моделі локально коригувати параметри апроксимації досліджуваної залежності. Роботу цього методу сформульовано такою послідовністю елементарних кроків:

*Крок 1.* Нормалізація вхідних векторів  $x_i \rightarrow \acute{x}_i$ ,  $i = \overline{1, n}$  з використанням лінійного масштабування (мін–макс нормалізація).

*Крок 2.* Задання меж для розбиття кожної ознаки  $\acute{x}_{i,j} \in [a, b]$  вектора  $\acute{x}_i$  на  $S$  частин:  $a = a_0 < a_1 < a_2 < \dots < a_{S-1} < a_S = b$ , де  $a$  – нижня,  $b$  – верхня межі нормалізованих даних.

*Крок 3.* Обчислення розширеного вектора локальних компонент  $\acute{x}_{i,j,s}$  для кожної  $j$ -тої ознаки  $\acute{x}_{i,j}$  вектора  $\acute{x}_i$  при  $s = \overline{1, S}$  за формулою:

$$\acute{x}_{i,j,s} = \begin{cases} 0, & \acute{x}_{i,j} \leq a_{s-1}, \\ \acute{x}_{i,j} - a_{s-1}, & a_{s-1} < \acute{x}_{i,j} \leq a_s, \\ a_s - a_{s-1}, & \acute{x}_{i,j} > a_s, \end{cases} \quad (3)$$

що гарантує збереження значення ознаки  $\acute{x}_{i,j} = \sum_{s=1}^S \acute{x}_{i,j,s}$ .

*Крок 4.* Об'єднання частин усіх розширених ознак згідно (3) у вектор  $x_i^+ \in \mathbb{R}^{mS}$  та формування розширеної матриці  $A^+ \in \mathbb{R}^{n \times mS}$  для подальшого навчання лінійної неітеративної нейроподібної структури МППП.

На етапі використання розвинутого методу необхідно виконати кроки 1-4 лише для вхідного вектора даних з невідомим значенням залежного атрибута. Кількість частин, на які розбиватиметься кожна компонента вхідного вектора, а також межі для цих частин мають бути такими ж, як і для навчальної вибірки даних.

Розвинутий метод дозволяє розбивати складну нелінійну залежність на сукупність локальних лінійних моделей, які легко оптимізувати, аналізувати та інтерпретувати. Крім того, використання лінійних неітеративних нейроподібних структур МППП для аналізу розширеної таким чином вибірки даних забезпечує швидкий процес навчання з гарантованою повторюваністю результату. Це особливо актуально при розв'язанні прикладних задач у біомедичних дослідженнях, де важлива не лише точність, але й

інтерпретованість результатів роботи моделі та її здатність до узагальнення.

Результати другого розділу висвітлено у наукових публікаціях [1, 7, 8, 9, 10, 28-33, 35].

**Третій розділ** дисертації присвячено розробці методології аугментації (збільшення обсягу) екстремально-короткої вибірки даних. Автором запропоновано принципово новий підхід до аугментації табличних вибірок даних. Такий підхід не використовує штучних методів введення додаткової інформації в дані шляхом інтерполяції, додавання шуму, генерації синтетичних даних тощо. Він базується на використанні інформації, наявної тільки у заданій вибірці даних, де ефект забезпечується виключно за допомогою операції декартового квадрата множини вхідних компонент векторів.

Декартовий квадрат вибірки даних  $A \in \mathbb{R}^{n \times m}$  подається як розширена матриця  $A^+ \in \mathbb{R}^{n^2 \times 2m}$ , утворена множиною всіх можливих впорядкованих пар векторів ознак  $x_q^+ = (x_i, x_k) \in \mathbb{R}^{2m}$ ,  $q = (i-1)n + k$ ,  $i, k = \overline{1, n}$ :

$$A^+ = \begin{bmatrix} (x_1^+)^T \\ (x_2^+)^T \\ \dots \\ (x_{n^2}^+)^T \end{bmatrix} \in \mathbb{R}^{n^2 \times 2m}, \quad (4)$$

де кожен рядок містить конкатенацію двох  $m$ -вимірних векторів  $(x_i, x_k)$ . Таким чином, аугментована матриця  $A^+$  має  $n^2$  рядків та  $2m$  стовпців.

Залежний атрибут  $z \in \mathbb{R}^{n^2 \times 1}$  для кожної такої пари формується як різниця залежних атрибутів обох векторів пари:

$$z_{(i-1)n+k} = y_k - y_i, \quad i, k = \overline{1, n}. \quad (5)$$

По суті (4) – це квадратичне збільшення обсягу вибірки даних для реалізації методів навчання ШНМ порівняно з початковою вибіркою даних. Окрім цього, вимірність кожного нового вектора подвоюється. Завданням такої непрямой постановки задачі регресії є апроксимація не значень залежних атрибутів досліджуваної вибірки даних, а різниць між ними, з подальшим використанням нелінійних ШНМ для реалізації процесу навчання.

Розглянемо ряд методів, поданих в межах розробленої методології. Базовий метод розв'язання задач регресії та класифікації, побудований на основі декартового квадрата множини вхідних компонент векторів, містить етапи навчання і застосування.

Етап навчання складається з послідовності таких кроків:

*Крок 1.* Аугментація для вибірки даних малого обсягу  $A$  згідно (4) та формування відповідних залежних атрибутів згідно (5).

*Крок 2.* Навчання нелінійної ШНМ на аугментованій вибірці даних  $A^+$  з першого кроку.

Для опису основних кроків процедури застосування базового методу введемо до розгляду вектор  $u \in \mathbb{R}^m$ , для якого невідоме значення залежного атрибута. Тоді етап застосування базового методу передбачає виконання таких

кроків:

*Крок 1.* Формування розширеної вибірки  $B^+ \in \mathbb{R}^{n \times 2m}$  шляхом конкатенації вектора  $u$  з кожним  $i$ -тим вектором  $x_i$  початкової вибірки даних  $A$ .

*Крок 2.* Отримання результатів  $z_i^{pred}$  шляхом застосування попередньо навченої нелінійної ШНМ до кожного розширеного вектора ознак, що входить до складу матриці  $B^+$ .

*Крок 3.* Агрегування результату:

- для задачі регресії:

$$y_u^{pred} = \frac{1}{n} \sum_{i=1}^n z_i^{pred} + \frac{1}{n} \sum_{i=1}^n y_i. \quad (6)$$

Формула (6) реалізує принцип часткової компенсації похибок через усереднення, властиві ансамблевим методам,

- для задачі класифікації:

$$y_u^{pred} = \arg \max_{i=1, n} (y_i + z_i^{pred}), \quad (7)$$

що відповідає принципу голосування більшістю, також характерному для ансамблевих методів.

Базовий метод забезпечує квадратичне збільшення обсягу вибірки без додавання шуму чи синтетичних векторів. Це підвищує ефективність навчання нелінійних моделей ШНМ у біомедичних задачах. Його ключова перевага – це навчання на відносних відмінностях між спостереженнями. Це дозволяє виявляти приховані залежності між об'єктами. Окрім цього, завдяки принципам усереднення або голосування більшістю реалізується принцип ансамблювання. Він зменшує вплив випадкових похибок та підвищує робастність методу до шуму у даних.

В адитивному методі розв'язання задачі на основі декартового квадрата множини вхідних компонент векторів, аналогічно до базового методу, використовується та сама процедура аугментації даних та навчання на нелінійних ШНМ. Однак ключовою особливістю цього методу є реалізація часткової взаємної компенсації випадкових похибок з протилежними знаками. Це досягається за рахунок симетричного формування значення залежного атрибута з (5) для кожної пари на етапі застосування методу і підвищує стійкість результату до шумів і флуктуацій у даних:

$$z_i^{(1)} = y_k - y_i, \quad z_i^{(2)} = y_i - y_k, \quad (8)$$

де значення  $k$  фіксоване і відповідає позиції вектора з невідомим значенням залежного атрибута,  $i$  – індекс усіх векторів навчальної вибірки даних  $i = \overline{1, n}$ .

Етап застосування адитивного методу вимагає послідовного виконання таких кроків:

*Крок 1.* Формування двох розширених вибірок даних  $B^{(1)+} \in \mathbb{R}^{n \times 2m}$  та  $B^{(2)+} \in \mathbb{R}^{n \times 2m}$  шляхом конкатенації вектора  $u$  з кожним  $i$ -тим вектором  $x_i$  з

А. Основна різниця – це позиція обох векторів  $u$  та  $x_i$  у розширеному векторі:  $u_i^{(1)+} = (u, x_i) \in \mathbb{R}^{2m}, i = \overline{1, n}; u_i^{(2)+} = (x_i, u) \in \mathbb{R}^{2m}, i = \overline{1, n}$ .

*Крок 2.* Отримання результатів  $z_i^{(1)pred}$  та  $z_i^{(2)pred}$  шляхом застосування попередньо навченої нелінійної ШНМ до кожного вектора матриць  $B^{(1)+}$  та  $B^{(2)+}$  відповідно.

*Крок 3.* Агрегування результату для розв'язання задачі регресії:

$$y_u^{pred} = \frac{\sum_{i=1}^n y_i}{n} + \frac{\sum_{i=1}^n z_i^{(1)pred}}{2n} - \frac{\sum_{i=1}^n z_i^{(2)pred}}{2n}. \quad (9)$$

Адитивний метод реалізує принцип ансамблювання через усереднення результату і парну компенсацію похибок. Це підвищує його робастність у випадку значного рівня випадкових флуктуацій у даних. Завдяки симетричному врахуванню взаємодій між вектором з невідомим виходом і всіма зразками навчальної вибірки досягається зменшення впливу шуму та покращення точності в умовах обмеженої кількості спостережень. Окрім цього, метод є гнучким до вибору базової моделі для реалізації процедури навчання. Це дозволяє інтегрувати як ядрові методи, так і нелінійні ШНМ та робить його придатним для застосування у задачах із значною вимірністю і малим обсягом даних, типових для біомедичних досліджень.

У деяких випадках інтелектуального аналізу екстремально-коротких вибірок даних використання відомих нелінійних ШНМ може демонструвати відносно високі результати. Проте отримана точність розв'язання конкретної задачі все ще може не відповідати вимогам користувача в певній області застосування. Для подолання цього обмеження доцільним є застосування принципів каскадування методів машинного навчання або формування комбінованої поверхні відгуку, які описано у другому розділі дисертації. Такі підходи дозволяють підвищити точність базових методів розв'язання задач регресії чи класифікації на основі декартового квадрата множини вхідних компонент векторів.

Для реалізації зазначених підходів автором розроблено відповідні методи. В них використовується додаткова інформація, сформована на першому етапі обробки, у вигляді нових ознак для аналізу. Зокрема, у випадку використання принципів каскадування ШНМ кожен вектор початкової вибірки даних  $A$  з (1) розширюється результатами, отриманими з попередньо навченої нелінійної ШНМ, яка забезпечує відносно високу точність результату. У методі комбінованої поверхні відгуку до кожного вектора додається лише один додатковий атрибут: залежне значення центра кластера, до якого належить відповідне спостереження. Далі для обох методів виконується операція декартового квадрата згідно з (4). Після цього реалізуються усі кроки базового або адитивного методу розв'язання задач регресії чи класифікації на основі декартового квадрата множини вхідних компонент векторів.

Розроблені каскадний метод і метод на основі формування комбінованої поверхні відгуку дозволяють суттєво підвищити точність розв'язання задач

регресії та класифікації, особливо в умовах екстремально коротких вибірок даних. У першому випадку розширення ознак відбувається за рахунок вихідних сигналів базової моделі ШНМ-1. Це суттєво покращує інформативність і знижує чутливість до шуму, забезпечуючи гнучкість у виборі моделей другого рівня та баланс між точністю і швидкодією. У другому – збільшення вимірності простору ознак виконується через включення інформації про середнє значення залежного атрибута в межах кластера. Це дозволяє враховувати геометричну структуру даних, згладжувати шум і компенсувати вплив аномалій. Обидва методи реалізують різні механізми підвищення властивості узагальнення методів, знижують ризик перенавчання та є ефективними для аналізу біомедичних даних різної вимірності в рамках розробленої методології.

Усі розроблені методи в цьому розділі базуються на авторській процедурі аугментації даних. Вона квадратично збільшує обсяг даних, що є перевагою при аналізі екстремально коротких вибірок даних. Проте для вибірок даних більшого обсягу це зумовлює суттєві обчислювальні витрати. Для подолання цього недоліку розроблено метод аугментації даних із використанням кластеризації. Він дозволяє суттєво зменшити кількість векторів для навчання без втрати, а іноді й з покращенням точності, завдяки збереженню репрезентативності вибірки через врахування її внутрішньої структури.

Етапи навчання та застосування методу формування аугментованої вибірки даних необхідного обсягу за рахунок кластеризації дещо відрізняються від вищенаведених.

Етап навчання цього методу реалізується через такі кроки:

*Крок 1.* Кластеризація початкової вибірки  $A \in \mathbb{R}^{n \times m}$  з (1) та визначення оптимальної кількості  $C_{opt}$  кластерів.

*Крок 2.* Формування матриці  $C \in \mathbb{R}^{C_{opt} \times (m+1)}$ , яка міститиме обчислені вектори центрів кластерів  $\mu_c \in \mathbb{R}^m$ ,  $c = \overline{1, C_{opt}}$  та значення їх залежних атрибутів  $y_c^{out} \in \mathbb{R}$ .

*Крок 3.* Аугментація вибірки незалежних ознак шляхом конкатенації кожного вектора ознак  $x_i$  початкової вибірки даних  $A$  з кожним центром кластеру  $\mu_c$  матриці  $C$ .

*Крок 4.* Пошук залежних атрибутів аугментованої вибірки даних для кожної відповідної пари  $(x_i, \mu_c)$ , наприклад, відповідно до базового методу (5).

*Крок 5.* Навчання нелінійної ШНМ на аугментованій вибірці даних.

Етап застосування методу формування аугментованої вибірки даних необхідного обсягу за рахунок кластеризації полягає у послідовному виконанні таких кроків:

*Крок 1.* Побудова матриці  $\tilde{B}^+ \in \mathbb{R}^{n \times 2m}$  шляхом конкатенації вектора  $u$  з кожним  $c$ -тим центром кластера  $\mu_c$  у випадку використання базового методу.

*Крок 2.* Отримання результатів  $\tilde{z}_c^{pred}$  шляхом застосування попередньо навченої нелінійної ШНМ до кожного вектора матриці  $\tilde{B}^+$ .

*Крок 3. Агрегування результату:*

- для задачі регресії:

$$y_u^{pred} = \frac{1}{C_{opt}} \sum_{c=1}^{C_{opt}} \tilde{z}_c^{pred} + \frac{1}{C_{opt}} \sum_{c=1}^{C_{opt}} y_c^{out}, \quad (10)$$

- для задачі класифікації:

$$y_u^{pred} = \arg \max_{c=1, C_{opt}} (y_c^{out} + \tilde{z}_c^{pred}). \quad (11)$$

На відміну від вищенаведених методів, розроблених у третьому розділі роботи, які характеризуються квадратичним збільшення обсягу вибірки даних  $n^2$ , для реалізації процедури навчання розроблений метод використовує центри кластерів  $C_{opt}$  для формування аугментованої вибірки суттєво меншого обсягу  $n \times C_{opt}$ , де  $n \times C_{opt} \ll n^2$ . Окрім цього, цей метод можна поєднувати з будь-яким іншим методом моделювання на основі декартового квадрата. Це забезпечить скорочення часу навчання при збереженні високої точності результатів його роботи.

Результати третього розділу висвітлено у наукових публікаціях [11-15, 26, 36-41].

**Четвертий розділ** дисертаційної роботи присвячено розробленню методів синтезу ансамблів ШНМ для розв'язання задач регресії та класифікації в умовах вибірок даних малих обсягів. В їх основу покладено комбінування глобально апроксимованих поверхонь відгуку. Розроблено методи синтезу ансамблів однорідних нейронних мереж лінійного або нелінійного типу, які дозволяють комбінувати як зміщені функції, так і зони кусково-лінійної апроксимації. Показано, що синтезовані ансамблі можуть бути реалізовані на основі усереднення чи голосування більшістю (клас бегінгових методів), а також блендингу (спрощеної схеми стекінгу) чи класичної схеми стекінгу.

Зокрема, метод синтезу ансамблю лінійних неітеративних нейроподібних структур МППП з  $L$  членів передбачає, що кожен з них навчається на власній унікальній вибірці даних, отриманій шляхом ізоморфного розширення вхідних ознак методом кусково-лінійної апроксимації багатовимірної поверхні відгуку.

Метод неітеративного навчання цього ансамблю включає кроки:

*Крок 1.* Нормалізація вибірки даних  $A \in \mathbb{R}^{n \times m}$  з (1) до діапазону  $[a, b]$  із використанням лінійного масштабування (мін–макс нормалізація).

*Крок 2.* Формування  $L$  випадкових схем розбиття кожної ознаки нормалізованого вектора даних  $\hat{x}_{i,j}$  на  $S$  частин шляхом випадкового задання меж цих частин для кожного  $l$ -того випадку за умови:

$$a = a_0^{(l)} < a_1^{(l)} < a_2^{(l)} < \dots < a_{S-1}^{(l)} < a_S^{(l)} = b, \quad l = \overline{1, L}. \quad (12)$$

*Крок 3.* Побудова унікальної розширеної вибірки даних для кожного з  $l$  членів ансамблю шляхом формування розширених векторів, компоненти яких обчислюються згідно (3):

$$\dot{x}_{i,j,s}^{(l)} = \begin{cases} 0, & \dot{x}_{i,j} \leq a_{s-1}^{(l)}; \\ \dot{x}_{i,j} - a_{s-1}^{(l)}, & a_{s-1}^{(l)} < \dot{x}_{i,j} \leq a_s^{(l)}, s = \overline{1, S}; \\ a_s^{(l)} - a_{s-1}^{(l)}, & \dot{x}_{i,j} > a_s^{(l)} \end{cases} \quad (13)$$

за умови збереження значення початкової ознаки  $\dot{x}_{i,j} = \sum_{s=1}^S \dot{x}_{i,j,s}^{(l)}$ .

*Крок 4.* Навчання кожного із  $L$  членів ансамблю на унікальній утвореній на попередньому кроці вибірці даних.

*Крок 5.* Агрегування результатів, яке може виконуватися шляхом усереднення вихідних сигналів усіх моделей ансамблю (бегінгова схема) або через навчання мета-моделі на їх виходах (стекінговий або блендінговий підхід).

На етапі застосування ансамблю виконуються вищенаведені кроки лише для одного вектора з невідомим значенням залежного атрибута.

Ансамбль лінійних неітеративних нейроподібних структур характеризується використанням єдиної модельної архітектури для всіх його елементів. При цьому різноманіття досягається за рахунок варіативності вхідних вибірок даних, сформованих методом кусково-лінійної апроксимації. Неітеративна природа навчання забезпечує обчислювальну ефективність та знижує ризик перенавчання, тоді як універсальність його побудови дозволяє адаптувати ансамбль до широкого класу задач регресії й класифікації з гарантованою повторюваністю результатів.

Метод синтезу ансамблів однорідних інтерполяційних ШНМ (НМУР або ЙНМ) ґрунтується на застосуванні випадкових зміщень для формування незалежних вибірок даних, унікальних для кожного члена ансамблю. В цьому випадку до всіх нормалізованих векторів ознак  $x_i \in \mathbb{R}^m$ ,  $i = \overline{1, n}$  вибірки даних  $A$  додається випадковий вектор зміщень  $\Delta = (\Delta_j) \in \mathbb{R}^m$ ,  $\Delta_j \in [-\delta, \delta]$ ,  $j = \overline{1, m}$ :

$$\tilde{x}_{ij} = x'_{ij} + \Delta_j, \quad j = \overline{1, m}. \quad (14)$$

Формується унікальна вибірка  $\tilde{A} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n]^\top \in \mathbb{R}^{n \times m}$ , яка зберігає обсяг та вимірність початкової, проте забезпечує різноманіття моделей ансамблю завдяки незалежності доданих зміщень. Це знижує кореляцію між їх вихідними сигналами і забезпечує необхідне різноманіття для ансамблю, сформованого з однотипних моделей ШНМ.

Метод навчання цього ансамблю вимагає послідовного виконання таких кроків:

*Крок 1.* Нормалізація вибірки даних  $\hat{A} = \mathcal{N}(A) = [\hat{x}_{i,j}], i = \overline{1, n}, j = \overline{1, m}$ .

*Крок 2.* Визначення оптимального параметра розмаху гаусівської функції для усіх НМУР чи ЙНМ ансамблю шляхом використання кросвалідації на початковій вибірці даних  $\hat{A}$ .

*Крок 3.* Генерація  $l$ -ої декорельованої вибірки даних  $\tilde{A}^{(l)}$  шляхом генерації  $l$ -того вектора зміщень  $\Delta^{(l)} = (\Delta_j^{(l)}) \in \mathbb{R}^m$  на основі (10) для  $l$ -того члена ансамблю  $l = \overline{1, L}$ .

*Крок 4.* Побудова ансамблю з  $L$  елементів, кожен з яких працює з власною, унікальною  $\tilde{A}^{(l)}$ -тою вибіркою даних, але з однаковим параметром розмаху гаусівської функції, визначеним на кроці 2.

*Крок 5.* Агрегування результатів усіх елементів ансамблю, наприклад, шляхом усереднення їх вихідних сигналів чи голосування більшістю (бегінгова схема).

На етапі застосування ансамблю виконується нормалізація вхідного вектора з невідомим значенням залежного атрибута. Далі цей вектор послідовно подається на кожен із  $l$  членів ансамблю, а отримані результати агрегуються згідно з обраною на етапі навчання схемою.

Ансамбль однорідних нелінійних ШНМ, сформований на основі випадкових зміщень, вирізняється простотою реалізації, використанням усієї доступної вибірки даних та високою ефективністю при аналізі даних малих обсягів, зокрема, завдяки використанню НМУР або ЙНМ. Наявність лише одного параметра для налаштування всіх членів ансамблю спрощує його конфігурацію та підвищує відтворюваність результатів.

Метод синтезу ансамблю неітеративних нейроподібних структур МПГП з розширенням входів на основі радіально-базисних функцій базується на використанні попарних симетричних випадкових зміщень для формування незалежних вибірок даних.

У цьому випадку для кожного нормалізованого вхідного вектора ознак  $x_i \in \mathbb{R}^m$ ,  $i = \overline{1, n}$  генерується вектор випадкових зміщень  $\Delta_j \in [-\delta, \delta]$ ,  $j = \overline{1, m}$  та його симетричний аналог  $-\Delta$ . Далі формуються два нові вектори  $\tilde{x}_{i,j}^{(+)} = x'_{ij} + \Delta_j$ ,  $\tilde{x}_{i,j}^{(-)} = x'_{ij} - \Delta_j$ ,  $j = \overline{1, m}$ . На їх основі будуються дві вибірки  $\tilde{A}^{(+)} \in \mathbb{R}^{n \times m}$  та  $\tilde{A}^{(-)} \in \mathbb{R}^{n \times m}$ , які об'єднуються у фінальну вибірку даних подвоєного обсягу:

$$\tilde{A}_{aug} = \begin{bmatrix} \tilde{A}^{(+)} \\ \tilde{A}^{(-)} \end{bmatrix} \in \mathbb{R}^{2n \times m}. \quad (15)$$

Це забезпечує утворення унікальної вибірки даних, відмінної від початкової. Повторна генерація різних векторів випадкових зміщень дозволяє створювати множину таких вибірок. Це уможливорює побудову ансамблю однорідних ШНМ із довільною кількістю елементів.

Метод неітеративного навчання цього ансамблю складається з таких кроків:

*Крок 1.* Нормалізація вибірки даних  $\hat{A} = \mathcal{N}(A) = [\hat{x}_{i,j}]$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$  та поділ її на дві підвибірки  $\hat{A} = \hat{A}_1 \cup \hat{A}_2$ , де  $\hat{A}_1$  призначена для навчання базових моделей (неітеративних нейроподібних структур МПГП з РБФ-розширенням входів), а  $\hat{A}_2$  – для навчання обраної мета-моделі.

*Крок 2.* Визначення оптимальних параметрів нейроподібної структури МПГП з РБФ-розширенням входів (кількості центрів РБФ та розмаху гаусівської функції) шляхом використання кросвалідації на  $\hat{A}_1$ .

*Крок 3.* Генерація  $l$ -тої аугментованої у два рази підвибірки даних  $\tilde{A}_1^{(l)} \in \mathbb{R}^{2n \times m}$  на основі використання  $\hat{A}_1$  та попарних симетричних відхилень (10) для  $l$ -того члена ансамблю  $l = \overline{1, L}$ .

*Крок 4.* Побудова ансамблю з  $L$  елементів, кожен з яких працює з власною, унікальною  $\tilde{A}^{(l)}$ -тою вибіркою даних, але з однаковими параметрами, визначеними на кроці 2.

*Крок 5.* Навчання усіх  $L$  базових елементів ансамблю.

*Крок 6.* Застосування  $\hat{A}_2$  на усіх попередньо навчених  $L$  базових елементів ансамблю та формування відповідних вихідних сигналів  $y_i^{pred(l)}$ .

*Крок 7.* Формування нової вибірки даних  $M = [y_i^{pred(1)}, \dots, y_i^{pred(L)}]$ ,  $M \in \mathbb{R}^{n \times L}$  і навчання мета-моделі.

На етапі застосування ансамблю вхідний вектор з невідомим значенням залежного атрибута спершу нормалізується. Далі він опрацьовується кожним із  $L$  членів ансамблю. Отримані вихідні сигнали формують розширений вектор ознак, який подається на мета-модель для обчислення остаточного результату.

У порівнянні з попередніми ансамблями розроблений на попарних симетричних зміщеннях метод має низку переваг. Завдяки нелінійному РБФ-розширенню ознак він краще виявляє складні локальні залежності та має кращі властивості узагальнення, ніж ансамбль лінійних нейроподібних структур МПГП. Метод попарних симетричних випадкових зміщень забезпечує формування унікальних вибірок, а також подвоює обсяг даних без втрати їх інформативності. Це підвищує як точність, так і робастність до шумів і викидів у даних завдяки компенсації похибок, зумовлених симетричними зміщеннями з протилежними знаками. Крім того, гнучкість параметризації ансамблю забезпечує баланс між складністю й властивостями узагальнення.

Результати четвертого розділу висвітлено у наукових публікаціях [2, 4-6, 42-47].

**П'ятий розділ** дисертації присвячено розробленню нових методів компенсації похибок ШНМ на основі каскадних ансамблів ШНМ.

Аналітично обґрунтовано можливість покращення результатів апроксимації та часткового усунення систематичної похибки нейронної мережі узагальненої регресії на основі використання каскаду з двох нейронних мереж. На основі цього розроблено метод компенсації похибки НМУР, етап навчання якого містить такі кроки:

*Крок 1.* Нормалізація вибірки даних  $\hat{A} = \mathcal{N}(A) = [\hat{x}_{i,j}], i = \overline{1, n}, j = \overline{1, m}$ .

*Крок 2.* Виконання основних кроків функціонування НМУР-1 ансамблю для опорної вибірки даних: пошук евклідових відстаней між поточним вектором  $\hat{x}_{k,j}$  та усіма іншими векторами  $\hat{x}_{i,j}$  з відомими вихідними значеннями  $y_i$  при ( $i \neq k$ ); обчислення гаусівських функцій на основі знайдених відстаней; визначення шуканого значення за формулою:

$$y_k^{pred} = \frac{\sum_{i \neq k}^n y_i \exp\left(-\frac{\sum_{j=1}^m (\dot{x}_{k,j} - \dot{x}_{i,j})^2}{2\sigma^2}\right)}{\sum_{i \neq k}^n \exp\left(-\frac{\sum_{j=1}^m (\dot{x}_{k,j} - \dot{x}_{i,j})^2}{2\sigma^2}\right)}, \quad (16)$$

де  $\sigma$  – розмах гаусівської функції.

*Крок 3.* Обчислення похибок роботи НМУР-1 для кожного  $i$ -го опорного вектора з відомим значенням вихідного сигналу:

$$\varepsilon_i = y_i - y_i^{pred}. \quad (17)$$

*Крок 4.* Агрегування результатів роботи ансамблю одним із трьох способів: шляхом простого підсумовування базового прогнозу  $y_k^{pred}$  з (16) та відповідної похибки  $\varepsilon_i$  з (17); за допомогою навчання коректуючої лінійної нейроподібної структури МППП, на входи якої подаються пари  $(y_k^{pred}, \varepsilon_i)$ , а на вихід – значення залежного атрибута  $y_i$ ; шляхом попереднього нелінійного збільшення вимірності простору вхідних ознак за допомогою полінома Колмогорова-Габора з подальшою корекцією результату лінійною неітеративною нейроподібною структурою МППП.

Етап застосування розробленого методу вимагає послідовного виконання таких кроків:

*Крок 1.* Нормалізація вхідного вектора  $\dot{u} = (\dot{u}_1, \dot{u}_2, \dots, \dot{u}_m)$  з невідомим значенням залежного атрибута.

*Крок 2.* Застосування НМУР-1 для пошуку базового результату:

$$y_u^{pred} = \frac{\sum_{i=1}^n y_i \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma^2}\right)}. \quad (18)$$

*Крок 3.* Застосування НМУР-2 для отримання похибки роботи НМУР-1 першого рівня ансамблю:

$$\varepsilon_u^{pred} = \frac{\sum_{i=1}^n \varepsilon_i \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma_{error}^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma_{error}^2}\right)}, \quad (19)$$

де  $\sigma_{error} \ll \sigma$ , а  $\varepsilon_i$  визначається з (17).

*Крок 4.* Агрегування результатів роботи НМУР-1 та НМУР-2 ансамблю одним із обраних на етапі навчання способів: підсумовуванням виходів НМУР-1 з (18) та НМУР-2 з (19); зваженим підсумовуванням їх результатів з використанням коректуючої лінійної неітеративної нейроподібної структури; нелінійним розширенням вихідних сигналів обох нейронних мереж за

допомогою полінома Колмогорова-Габора з подальшим коригуванням результатів додатковою лінійною ШНМ.

Розроблений метод підвищує точність аналізу даних малих обсягів на основі нейронної мережі узагальненої регресії шляхом часткового усунення її систематичної похибки. Різні варіанти агрегування результатів забезпечують гнучкість цього методу: просте підсумовування базового результату та оціненої похибки не потребує додаткового навчання і є ефективним для задач із лінійним характером залишкових похибок; застосування коректуючої лінійної нейроподібної структури МППП забезпечує точніше відтворення заданої залежності навіть за умов екстремально короткої вибірки даних; використання поліноміального розширення входів додатково враховує складні нелінійні взаємозв'язки, покращуючи властивості узагальнення у випадках нелінійної залежності між похибкою і базовим результатом.

Автором також вдосконалено метод глобально-локальної апроксимації, який на відміну від існуючого забезпечує незалежне виділення обох компонент сигналу за допомогою ансамблю з двох НМУР та дозволяє реалізувати ітеративне уточнення шуканого значення. Етап навчання удосконаленого методу включає такі кроки:

*Крок 1.* Формування розширеної навчальної вибірки даних  $A^+ = [x_1^+, \dots, x_n^+]^T$ ,  $A^+ \in \mathbb{R}^{n \times (m+1)}$  з додаванням залежного атрибута  $y_i$  у простір ознак  $x_i^+ = (x_i, y_i) \in \mathbb{R}^{m+1}$ ,  $i = \overline{1, n}$ , її нормалізація та навчання лінійної нейроподібної структури МППП зі звуженим горлом для виділення глобальної компоненти поверхні відгуку  $y_i^{global}$ .

*Крок 2.* Обчислення локальної компоненти поверхні відгуку для навчальної вибірки даних:

$$y_i^{local} = y_i - y_i^{global}. \quad (20)$$

Етап застосування удосконаленого методу реалізується через послідовність таких кроків:

*Крок 1.* Отримання попередньої оцінки глобальної компоненти із використанням першої нейронної мережі узагальненої регресії (НМУР-1):

$$y_u^{(0)} = \frac{\sum_{i=1}^n y_i \exp\left(-\frac{\sum_{j=1}^m (\hat{u}_j - \hat{x}_{i,j})^2}{2\sigma_{glob}^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{\sum_{j=1}^m (\hat{u}_j - \hat{x}_{i,j})^2}{2\sigma_{glob}^2}\right)}, \quad (21)$$

де  $\sigma_{glob}$  — розмах гаусівської функції для НМУР-1.

*Крок 2.* Побудова розширеного вектора  $u^+ = (\hat{u}, \hat{y}_u^{(0)}) \in \mathbb{R}^{m+1}$  з подальшим застосуванням попередньо навченої лінійної нейроподібної структури для більш точного виділення глобальної компоненти поверхні відгуку  $y_u^{global}$ .

*Крок 3.* Оцінка локальної компоненти поверхні відгуку для  $u$  із використанням другої нейронної мережі узагальненої регресії (НМУР-2):

$$y_u^{local} = \frac{\sum_{i=1}^n (y_i - y_i^{global}) \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma_{loc}^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{\sum_{j=1}^m (\dot{u}_j - \dot{x}_{i,j})^2}{2\sigma_{loc}^2}\right)}, \quad (22)$$

де  $\sigma_{loc}$  — розмах гаусівської функції для НМУР-2, причому  $\sigma_{loc} \ll \sigma_{glob}$ .

*Крок 4.* Агрегування результатів роботи удосконаленого методу глобально локальної апроксимації поверхні відгуку для вектора  $u$ :

$$y_u^{res} = y_u^{global} + y_u^{local}. \quad (23)$$

Для підвищення точності (23) значення  $y_u^{res}$  можна повторно використати у (20). Це дозволяє реалізувати ітеративне уточнення результату до досягнення заданого критерію зупинки.

Удосконалений метод забезпечує підвищену точність апроксимації біомедичних вибірок даних малих обсягів за рахунок незалежного виділення глобальної та локальної компонент вихідного сигналу, де груба оцінка закономірності виконується НМУР-1, уточнена глобальна оцінка — лінійною неітеративною нейроподібною структурою МППП, а локальні залишки моделюються НМУР-2. Сукупно це дозволяє суттєво зменшити похибку без ускладнення процесу навчання.

У цьому розділі дисертації також розроблено прямий і зворотний методи компенсації похибки моделі регресії на основі використання раціональних дробів. Останні, в порівнянні з поліноміальними, забезпечують можливість ефективно відтворювати асимптотичні властивості та локальні особливості функції при відносно невисоких степенях чисельника й знаменника. Особливо це помітно, коли слід зберегти точність на великому діапазоні аргументів, що характерно для біомедичних досліджень.

Прямий метод часткового усунення похибки моделі регресії реалізується дворівневим каскадним ансамблем з можливістю ітеративного уточнення результату. Етап навчання прямого методу включає такі кроки:

*Крок 1.* Нормалізація вхідних даних  $\dot{x}_i = \mathcal{N}(x_i)$  та побудова базової моделі регресії  $f_1: \mathbb{R}^m \rightarrow \mathbb{R}$  (довільна ШНМ чи алгоритм машинного навчання), яка дає початкову оцінку:

$$y_i^{pred} = f_1(\dot{x}_i), i = \overline{1, n}. \quad (24)$$

*Крок 2.* Обчислення відносної похибки базової моделі для навчальної вибірки:

$$t_i = \frac{y_i^{pred}}{y_i} - 1, i = \overline{1, n}. \quad (25)$$

*Крок 3.* Навчання коректуючої моделі  $f_2(\cdot)$  на початковій навчальній вибірці вхідних даних, де вихід моделі задається як  $t_i$ :

$$t_i^{pred} = f_2(\dot{x}_i), i = \overline{1, n}. \quad (26)$$

*Крок 4.* Уточнення вихідного сигналу на навчальній вибірці:

$$y_i^{(1)} = \frac{y_i^{pred}}{t_i^{pred} + 1}. \quad (27)$$

Етап застосування прямого методу компенсації похибки моделі регресії вимагає послідовного виконання таких кроків:

*Крок 1.* Нормалізація вхідного вектора даних  $u$ ,  $\dot{u} = \mathcal{N}(u)$  та застосування попередньо навченої базової моделі  $f_1(\cdot)$  для отримання первинного результату:

$$y_u^{pred} = f_1(\dot{u}). \quad (28)$$

*Крок 2.* Прогноз відносної похибки базової моделі  $t_u^{pred}$ , використовуючи попередньо навчену коректуючу модель  $f_2(\cdot)$ :

$$t_u^{pred} = f_2(\dot{u}). \quad (29)$$

*Крок 3.* Уточнення результату:

$$y_u^{(1)} = \frac{y_u^{pred}}{t_u^{pred} + 1}. \quad (30)$$

За необхідності отримане значення  $y_u^{(1)}$  можна використати як нове  $y$  (25). Тоді наступні кроки повторюються ітеративно до досягнення заданого критерію зупинки для підвищення точності апроксимації.

Переваги цього методу полягають у тому, що він знижує відносну похибку без зміни базової моделі, підтримує ітеративне уточнення результату, потребує лише вхідних та вихідних даних (без знання внутрішньої структури  $f_1(\cdot)$ ) і демонструє високу ефективність розв'язання задач інтелектуального аналізу вибірок даних малого обсягу.

Зворотний метод часткового усунення похибки моделі регресії також реалізовано у вигляді дворівневого каскадного ансамблю з можливістю ітеративного уточнення результату. Проте тут є ряд відмінностей. Етап навчання зворотного методу містить такі кроки:

*Крок 1.* Нормалізація вхідних даних  $\dot{x}_i = \mathcal{N}(x_i)$  та побудова базової моделі регресії  $f_1: \mathbb{R}^m \rightarrow \mathbb{R}$  (довільна ШНМ чи алгоритм машинного навчання), яка дає початкову оцінку:

$$y_i^{pred} = f_1(\dot{x}_i), i = \overline{1, n}. \quad (31)$$

*Крок 2.* Обчислення значень залишкової функції:

$$r_i = \frac{y_i}{y_i^{pred}} - 1, i = \overline{1, n}. \quad (32)$$

*Крок 3.* Навчання коректуючої моделі  $f_3(\cdot)$  на початковій навчальній вибірці вхідних даних, де її вихід задається як  $r_i$ :

$$r_i^{pred} = f_3(\dot{x}_i), i = \overline{1, n}. \quad (33)$$

*Крок 4.* Уточнення вихідного сигналу на навчальній вибірці:

$$y_i^{(1)} = y_i^{pred} (1 + r_i^{pred}). \quad (34)$$

Етап застосування зворотного методу компенсації похибки моделі регресії включає такі кроки:

*Крок 1.* Нормалізація вхідного вектора даних  $u$ ,  $\dot{u} = \mathcal{N}(u)$  та застосування попередньо навченої базової моделі  $f_1(\cdot)$  для отримання первинного результату:

$$y_u^{pred} = f_1(\dot{u}). \quad (35)$$

*Крок 2.* Прогноз відносної корекції базової моделі  $r_u^{pred}$ , використовуючи попередньо навчену коректуючу модель  $f_3(\cdot)$ :

$$r_u^{pred} = f_3(\dot{u}). \quad (36)$$

*Крок 3.* Уточнення результату:

$$y_u^{(1)} = y_u^{pred} (1 + r_u^{pred}). \quad (37)$$

Можна також ітеративно замінити  $y_i^{pred}$  на  $y_u^{(1)}$  у (32) і повторити інші кроки до досягнення заданого критерію зупинки.

Порівняно з прямим методом  $\left(y_i \approx \frac{y_i^{pred}}{1+f_3(\dot{x}_i)}\right)$ , який коригує прогноз шляхом ділення на відносну похибку, зворотний метод  $\left(y_i \approx y_i^{pred} (1 + f_3(\dot{x}_i))\right)$  підсилює початковий результат множенням на  $(1 + f_3(\dot{x}_i))$ . Це забезпечує більш точне зменшення систематичних похибок без ризику надмірного коригування малих вихідних значень.

Результати п'ятого розділу висвітлено у наукових публікаціях [3, 16-22].

**Шостий розділ** присвячено перевірці правильності, функціональності та базової ефективності розроблених та удосконалених методів на основі застосування існуючих або синтетичних вибірок даних із наперед відомими властивостями. Для цього автором створено низку спеціалізованих програмних засобів мовою Python, які реалізують усі досліджувані методи та захищені свідоцтвами про авторське право на комп'ютерні програми.

Верифікацію методу нелінійного доповнення входів лінійної неітеративної нейроподібної структури з використанням полінома Колмогорова-Габора виконано на задачі оцінювання витрат на медичне страхування. Застосування полінома Колмогорова-Габора другого степеня суттєво підвищило точність порівняно з базовою моделлю, тоді як оптимізація архітектури мережі покращила властивості узагальнення та скоротила час навчання у більш, ніж 2,5 рази. Порівняльний аналіз засвідчив перевагу розробленого підходу над класичними лінійними регресорами за всіма показниками точності та забезпечив можливість аналітичного представлення

результатів у вигляді поліноміального виразу. Це підвищує інтерпретованість моделі та рівень довіри до її висновків з боку клінічних експертів.

Оцінювання ефективності удосконаленого методу класифікації даних на основі формування комбінованої поверхні відгуку з використанням кластерного аналізу здійснено на прикладі задачі прогнозування ризику серцевого нападу. Моделювання проведено з використанням вибірки даних, що включала 294 пацієнтів і 18 медичних ознак. Удосконалення архітектури двошарового ансамблю на базі лінійної неітеративної нейроподібної структури та застосування кластеризації дозволило скоротити кількість вхідних атрибутів. Це забезпечило покращення властивостей узагальнення моделі при скороченні часу навчання фінального класифікатора на 23 %. Найвищу точність класифікації досягнуто за умови застосування кластеризації з урахуванням вихідного атрибута та оптимальною кількістю кластерів, яка становила 7. У такій конфігурації метод продемонстрував точність 93,2 % у режимі застосування, що на 4,7 % перевищує показник існуючого методу.

Тестування базового методу розв'язання задачі регресії на основі декартового квадрата множини вхідних компонент векторів здійснено на штучній вибірці даних. Вона складалась із 30 спостережень, згенерованих на основі існуючої нелінійної функції. Процедура навчання методу виконувалася на основі використання неітеративної нейроподібної структури МППП з РБФ-розширенням входів. Отримано високу точність ( $R^2 = 0,99$ ) та мінімальну похибку ( $RMSE = 0,0023$ ) у режимі застосування. Порівняльний аналіз із класичними та сучасними методами показав, що розроблений метод забезпечує значно менше значення  $RMSE$ , ніж найближчий за точністю аналог, повністю уникаючи проблеми перенавчання при роботі з екстремально-короткими вибірками даних.

Для перевірки ефективності адитивного методу, побудованого на ідеї використання декартового квадрата, проведено моделювання для задачі з галузі ревматології. На вибірці з 35 спостережень метод забезпечив значення  $RMSE = 1,95$  і  $R^2 = 0,91$ . Це на 42 % точніше за найкращий із відомих методів. Крім того, експерименти засвідчили, що ітеративні нелінійні алгоритми машинного навчання також можуть бути використані для реалізації процедури навчання у розроблених методах. Водночас, хоча це й покращує точність, отримане її зростання є менш суттєвим, ніж у випадку з неітеративними нейроподібними структурами, тоді як час навчання істотно збільшується.

Розроблений метод класифікації на основі декартового квадрата множини вхідних компонент із використанням принципів каскадування ШНМ перевірено на задачі з галузі трансплантології. Класифікація результатів трансплантації стовбурових клітин у дітей на основі вибірки з 187 пацієнтів і 36 ознак засвідчила високу точність ( $F1 = 0,98$ ), що є найвищою серед усіх порівнюваних методів.

Доповненням до методів цього класу є підхід до формування аугментованої вибірки необхідного обсягу шляхом кластеризації. Він дає змогу суттєво скоротити обсяг навчальних даних (у 9 разів на тестових даних)

без втрати точності ( $R^2 = 0,996$ ). Цей метод є особливо ефективним у задачах, що вимагають швидкодії, та може бути інтегрований у будь-який із трьох вищевказаних методів без зниження їх властивостей узагальнення. Вибір між повною та скороченою формами аугментування залежить від вимог до точності, обсягу вихідних даних і допустимого часу навчання.

Перевірку ефективності для трьох методів часткової компенсації похибок моделей регресії з розділу 5, орієнтованих на підвищення точності в умовах вибірок даних малих обсягів, виконано на прикладі задач із трьох різних предметних галузей: екологічного моніторингу, біомедичної інформатики та біоматеріалознавства. Усі методи реалізовано на основі каскадних ансамблів із використанням неітеративних нейроподібних структур. Зокрема, розроблений метод компенсації похибок нейронної мережі узагальненої регресії протестовано на задачі відновлення втрачених вимірювань концентрації окису вуглецю CO у повітрі. Метод протестовано із використанням трьох схем агрегування результатів, які забезпечили зменшення похибки нейронної мережі узагальненої регресії в межах 4-8% на основі RMSE. Удосконалений метод глобально-локальної апроксимації застосовано до задачі оцінки вмісту жирової тканини за антропометричними параметрами. На вибірці зі 252 спостережень метод досяг значень  $RMSE = 3,81$  та  $R^2 = 0,84$ . Це на 13% точніше за існуючий метод та суттєво переважає інші моделі, включно з Випадковим Лісом та Машиною опорних векторів. Розроблений метод компенсації похибок довільної моделі регресії на основі раціональних дробів (прямий і зворотний підходи) використано для покращення точності регресора на основі Мащини опорних векторів у задачі оцінювання твердості біокераміки  $Al_2O_3$ . Застосування прямого і зворотного варіантів методу на основі неітеративних нейроподібних структур дозволило знизити RMSE від значення 1,53 до значення 1,48. Зворотний варіант показав кращу ефективність завдяки більш гнучкій адаптації до масштабів помилки.

Результати шостого розділу висвітлено у наукових публікаціях [8, 13, 23, 25, 28, 29, 42, 43, 51-53].

**Сьомий розділ** присвячено тестуванню та впровадженню низки розроблених методів для розв'язання різних прикладних задач.

Метод класифікації даних на основі розширення входів лінійної нейроподібної структури вихідними сигналами шару підсумовування ймовірнісної нейронної мережі впроваджено для розв'язання задачі в галузі біоматеріалознавства. Застосування цього методу забезпечило високі значення точності, влучності, повноти та F1-міри як на етапі навчання, так і в процесі практичного використання для класифікації матеріалів з порошкових титанових сплавів ще на стадії підготовки сировини. Це має критично важливе значення для виробництва біосумісних медичних імплантатів із заданими експлуатаційними властивостями. Результати впроваджено на кафедрі стоматології дитячого віку та імплантології Харківського національного медичного університету впродовж 2022–2024 років (акт впровадження від 04.03.2024 р.).

Метод кусково-лінійної апроксимації багатовимірної поверхні відгуку впроваджено для розв'язання задачі оцінювання біологічного віку трабекулярної кістки людини на основі структурно-функціональних характеристик тканини в умовах аналізу вибірки даних малого обсягу (для пацієнтів з клінічно діагностованим тяжким остеоартрозом). Застосування цього методу засвідчило високу точність результату (коефіцієнт детермінації  $R^2 = 0,92$ ) та забезпечило зменшення середньоквадратичної помилки майже на 58 % порівняно з лінійною нейроподібною структурою без попереднього розширення простору ознак. Такий підхід є важливим елементом сучасної стратегії ранньої діагностики дегенеративних змін кісткової тканини, зокрема, остеопорозу та остеоартрозу. Ефективне розв'язання цієї задачі дає змогу не тільки виявляти наявність патологічних змін, а й прогнозувати їх прогресування. Це принципово важливо для своєчасного клінічного втручання та корекції терапії. Метод впроваджено в практичну діяльність Комунального некомерційного підприємства Харківської обласної ради «Обласна клінічна лікарня» в період з жовтня 2022 року по листопад 2024 року (акт впровадження від 02.12.2024 р.).

Метод розв'язання задачі регресії на основі використання декартового квадрата множини вхідних компонент векторів з формуванням комбінованої поверхні відгуку впроваджено для розв'язання задач в галузі медичної стоматології. Його застосування для оцінювання механічних властивостей цирконієвої кераміки, зокрема, в'язкості руйнування (тріщиностійкості) продемонструвало високу точність результатів (коефіцієнт детермінації  $R^2 = 0,816$ ). Це суттєво перевищує показники, отримані з використанням класичних методів машинного навчання, а також забезпечує на 16 % вищу точність порівняно з відомим методом. Ефективне розв'язання цієї задачі забезпечує підвищення якості керамічних виробів ще на етапі спікання порошкової суміші на основі оксиду цирконію та подовження терміну їх експлуатації, а також дозволяє істотно зменшити витрати часу, фінансових і матеріальних ресурсів, пов'язаних із використанням високотехнологічного обладнання для експериментального визначення відповідної характеристики традиційними методами. Метод розроблено та реалізовано в межах виконання наукового гранту EURIZON: *Artificial Intelligence in Medicine: Small Data Approach (SmallAIM)* (№ 871072, 01.04.2024 – 31.03.2025), що фінансувався Європейською Комісією в рамках програми *Horizon Europe*. Результати впроваджено в практичну діяльність (акт впровадження № 17244/3 від 17.03.2025 р.).

Метод розв'язання задачі класифікації даних на основі ансамблю ймовірнісних нейронних мереж із використанням симетричних випадкових зміщень впроваджено для розв'язання задачі класифікації летальних випадків серед пацієнтів із політравмами на основі 56 клінічних і лабораторних показників. Його використання підтвердило високу ефективність в умовах екстремально короткої вибірки даних ( $n = 73$ ), зібраної у Харківській міській клінічній лікарні швидкої та невідкладної медичної допомоги імені професора

О. І. Мещанінова. Розроблений метод забезпечив найвищу точність класифікації серед усіх порівнюваних методів, зокрема, за F1-мірою (0,91) і коефіцієнтом кореляції Метьюса (0,90) у стекінговій конфігурації ансамблю. Отримані результати забезпечили додаткову інформацію лікареві для оцінювання ризику смерті, оптимізації тріажу в умовах інтенсивної терапії та моніторингу ефективності лікування. Розроблений метод впроваджено в межах діяльності Військово-медичного клінічного центру Західного регіону в період з січня 2024 року по січень 2025 року (акт впровадження від 22.04.2025).

Метод розв'язання задачі класифікації даних на основі ансамблю неітеративних нейроподібних структур із розширенням входів радіально-базисними функціями, у якому для декореляції вибірок даних застосовуються симетричні випадкові зміщення, впроваджено для задачі класифікації результату виживаності ниркового трансплантату впродовж першого місяця після операції з трансплантації на основі 40 клініко-лабораторних ознак. Його використання дозволило досягти високих показників точності (F1-міра = 0,91 у режимі застосування) та забезпечити найкращі результати серед усіх порівнюваних методів. Застосування методу є особливо важливим під час прийняття клінічних рішень у післяопераційний період, коли необхідне точне оцінювання ризиків відторгнення трансплантату, адаптація терапевтичних протоколів та індивідуалізація імуносупресивної підтримки. Розроблений метод впроваджено в лікувальну практику Комунального некомерційного підприємства Харківської обласної ради “Обласний медичний клінічний центр урології і нефрології ім. В. І. Шаповала” в період з жовтня 2022 року по вересень 2024 року (акт впровадження від 22.09.2024 р.).

Результати сьомого розділу висвітлено у наукових публікаціях [8, 13, 23, 25, 28, 29, 42, 43, 51-53].

## ВИСНОВКИ

У дисертаційній роботі на основі виконання теоретичних та експериментальних досліджень вирішено важливу теоретичну проблему розвитку основ теорії синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для підвищення точності розв'язання задач регресії та класифікації на вибірках біомедичних даних малих обсягів. При цьому отримано такі основні результати та висновки:

1. Проведено аналіз методів та засобів інтелектуального опрацювання структурованих біомедичних даних малого обсягу та обґрунтовано необхідність синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для підвищення точності і повторюваності результатів їх обробки.

2. Набув подальшого розвитку метод кусково-лінійної апроксимації для випадку аналізу багатовимірних поверхонь відгуку, який за рахунок ізоморфного подання кожного значення ознаки у вигляді її локальних

компонент дав можливість збільшити вимірність простору вхідних даних та забезпечити локальну корекцію параметрів поверхні відгуку лінійною неітеративною нейроподібною структурою МППП, що дозволяє зменшити похибку лінійної моделі на 58% згідно RMSE.

3. Удосконалено метод формування комбінованої поверхні відгуку з застосуванням кластерного аналізу, який на відміну від існуючого доповнює простір початкових ознак тільки значенням залежного атрибута центра кластера, що забезпечує мінімальне зростання вимірності простору ознак та підвищує точність моделей регресії і класифікації на малих вибірках без втрати здатності до узагальнення.

4. Розроблено метод нелінійного розширення простору незалежних ознак біомедичної вибірки даних, який внаслідок їх доповнення вихідними сигналами шару підсумовування ймовірнісної нейронної мережі дозволяє підвищити точність класифікації за рахунок використання лінійних неітеративних нейроподібних структур МППП на 43% та 9% згідно F1-міри у порівнянні з обома формуючими ШНМ при збереженні властивостей узагальнення.

5. Розроблена методологія розв'язання задач регресії та бінарної класифікації на екстремально-коротких вибірках біомедичних даних шляхом аугментації даних на основі декартового квадрата множини вхідних компонент векторів дозволяє застосовувати різнотипні нелінійні моделі машинного навчання, зокрема, неітеративні нейроподібні структури МППП з RBF-розширенням входів для реалізації навчання та забезпечує:

- покращення властивостей узагальнення та зниження фактору перенавчання базового методу завдяки непрямій постановці задачі регресії, яка дозволяє моделі змістити акценти на відносні відмінності між спостереженнями і підвищити точність розв'язання задач регресії нелінійними ШНМ;

- підвищення робастності до аномалій та викидів у даних з використанням адитивного методу за рахунок можливості часткової взаємної компенсації похибок з різними знаками при формуванні значення шуканого сигналу методом усереднення або голосування більшістю, а також зменшення похибки RMSE методу до 59 % в процесі розв'язання задач бінарної класифікації порівняно з методом опорних векторів;

- підвищену здатність до апроксимації суттєво нелінійних поверхонь відгуку з використанням каскадного методу за рахунок додаткового розширення вхідного простору початкової вибірки даних результатами роботи ймовірнісної нейронної мережі на першому етапі методу, а також підвищення точності базової ШНМ на 10 % відповідно до значень коефіцієнта кореляції Метьюса;

- можливість апроксимації суттєво нелінійної поверхні відгуку, заданої екстремально короткою вибіркою даних, шляхом формування комбінованої поверхні відгуку, що дозволяє точніше моделювати залежності з урахуванням локальної структури даних та підвищити адаптивність моделі до складних

випадків, зокрема, коли класи частково перекриваються або спостерігається значна варіативність значення залежних атрибутів;

- зменшення обсягу аугментованої вибірки даних базового методу майже у 9 разів при зниженні точності менше 1% за коефіцієнтом детермінації завдяки використанню методу формування аугментованої вибірки необхідного обсягу на основі кластеризації.

6. Розроблено метод синтезу ансамблю лінійних нейроподібних структур з неітеративним машинним навчанням, який завдяки агрегуванню кускових лінійно апроксимованих поверхонь відгуку кожного його елемента забезпечує підвищення точності розв'язання задачі регресії на 5% відповідно до значень коефіцієнта детермінації при збереженні високої швидкодії його функціонування та простоти реалізації.

7. Розроблено метод синтезу ансамблів неітеративних нейроподібних структур з РБФ-розширенням входів, який шляхом агрегування мета-моделлю апроксимованих поверхонь відгуку, сформованих на незалежних вибірках з використанням симетричних випадкових зміщень, забезпечує ефективну аугментацію, декореляцію та часткову компенсацію похибок з протилежними знаками, що підвищує точність розв'язання задачі регресії майже на 6 % відповідно до значень коефіцієнта кореляції Метьюса.

8. Розроблено метод часткового усунення систематичної похибки нейронної мережі узагальненої регресії, який за рахунок використання ансамблю з двох таких нейронних мереж та трьох способів агрегування їх результатів дозволяє зменшити похибки функціонування ШНМ на 4 – 8% відповідно до RMSE в процесі розв'язання задачі регресії у випадку аналізу біомедичних даних малих обсягів.

9. Удосконалено метод глобально локальної апроксимації біомедичних даних малих обсягів, який на відміну від існуючого методу виконує незалежне виділення глобальної та локальної компонент шуканого сигналу, використовуючи ансамбль з двох нейронних мереж узагальненої регресії з можливістю ітеративного уточнення результату. Це забезпечує зменшення похибки RMSE на 13% при розв'язанні задачі регресії в умовах аналізу біомедичних даних обмеженого обсягу.

10. Розроблено прямий та зворотний методи компенсації похибки моделі регресії, які завдяки використанню перетворень залежностей у вигляді раціональних дробів дозволяють застосовувати ітеративне уточнення результатів, підвищуючи точність довільної моделі регресії без потреби у знанні її внутрішньої структури або параметрів, забезпечуючи зменшення похибки MSE, наприклад, машини опорних векторів на 9% та 11% відповідно.

11. Розвинуто теорію синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням, яка ґрунтується на розроблених методах нелінійного розширення простору ознак, аугментації даних, декореляції вибірок та компенсації похибок, що дає можливість підвищити точність розв'язання задач регресії та класифікації на вибірках біомедичних даних малих обсягів.

12. Розроблені та захищені авторським правом на комп'ютерні програми програмні засоби забезпечують можливість експериментального оцінювання ефективності розроблених та вдосконалених методів як на модельних, так і на реальних даних біомедичних досліджень. Отримані результати апробовані при розв'язанні низки прикладних задач у різних напрямках біомедичних досліджень, зокрема, в біоматеріалознавстві, ревматології, травматології, стоматології та трансплантології. Їх впровадження виконано у співпраці з медичними закладами, науково-освітніми установами та в межах виконання науково-дослідних робіт в Україні та міжнародних грантових проєктів за кордоном. Це засвідчено відповідними актами впровадження та використання результатів досліджень. Етичні аспекти враховані як на етапі моделювання, так і в процесі практичного впровадження розроблених та удосконалених методів, що забезпечує відповідальне їх використання в біомедичних дослідженнях.

### СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

1. Izonin I. V. An unsupervised-supervised ensemble technology with non-iterative training algorithm for small biomedical data analysis // *Computer systems and information technologies*. 2023. No. 4. P. 67–74. DOI: 10.31891/csit-2023-4-9 (фахове видання України, категорія Б).
2. Tkachenko R. O., Izonin I. V., Danylyk V. M., Mykhalevych V. Yu. Stacking of the SGTm neural-like structure with RBF layer based on generation of a random curtain of its hyperparameters for prediction tasks // *Ukrainian Journal of Information Technology*. 2021. Vol. 3, No. 1. P. 49–55. DOI: 10.23939/ujit2021.03.049 (фахове видання України, категорія Б).
3. Izonin I. V. An ensemble method for the regression model parameter adjustments: direct approach // *Mathematical and computer modelling. Series: Technical sciences*. 2023. Vol. 24. P. 35–44. DOI: 10.32626/2308-5916.2023-24.35-44 (фахове видання України, категорія Б).
4. Вітинський П. В., Ткаченко Р. О., Ізонін І. В., Кустрі Н. О. Ансамблі нейроподібних структур МПГП з RBF розширенням входів для задач регресії та класифікації // *Вісник Хмельницького національного університету. Серія: Технічні науки*. 2019. № 4(275). С. 72–79. (фахове видання України, категорія Б).
5. Izonin I. V., Tkachenko R. O., Semchyshyn O. L. An ensemble method for the analysis of small biomedical data based on a neural network without training // *Elektronnoe modelirovanie*. 2023. Vol. 45, No. 6. P. 65–76. DOI: 10.15407/emodel.45.06.065 (фахове видання України, категорія Б).
6. Вітинський П. В., Ткаченко Р. О., Ізонін І. В. Ансамбль мереж GRNN для розв'язання задач регресії з підвищеною точністю // *Науковий*

*вісник НЛТУ України*. 2019. Т. 29, № 8. Стаття № 8. DOI: 10.36930/40290822 (фахове видання України, категорія Б).

7. Tkachenko R., Izonin I., Vitynskyi P., Lotoshynska N., Pavlyuk O. Development of the non-iterative supervised learning predictor based on the Ito decomposition and SGTM neural-like structure for managing medical insurance costs // *Data*. 2018. Vol. 3, No. 4. P. 46. DOI: 10.3390/data3040046. (Індексується у Scopus).

8. Izonin I., Tkachenko R., Gregus M., Duriagina Z., Shakhovska N. PNN-SVM approach of Ti-based powder's properties evaluation for biomedical implants production // *Computers, Materials & Continua*. 2022. Vol. 71, No. 3. P. 5933–5947. DOI: 10.32604/cmc.2022.022582. (Індексується у Scopus, квартиль Q2).

9. Izonin I., Tkachenko R., Shakhovska N., Ilchyshyn B., Singh K. K. A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain // *Mathematics*. 2022. Vol. 10, No. 11. P. 1942. DOI: 10.3390/math10111942. (Індексується у Scopus, квартиль Q2).

10. Izonin I., Tkachenko R., Duriagina Z., Shakhovska N., Kovtun V., Lotoshynska N. Smart web service of Ti-based alloy's quality evaluation for medical implants manufacturing // *Applied Sciences*. 2022. Vol. 12, No. 10. P. 5238. DOI: 10.3390/app12105238. (Індексується у Scopus, квартиль Q2).

11. Izonin I., Tkachenko R., Dronyuk I., Tkachenko P., Gregus M., Rashkevych M. Predictive modeling based on small data in clinical medicine: RBF-based additive input-doubling method // *Mathematical Biosciences and Engineering*. 2021. Vol. 18, No. 3. P. 2599–2613. DOI: 10.3934/mbe.2021132. (Індексується у Scopus, квартиль Q2).

12. Izonin I., Tkachenko R., Shakhovska N., Lotoshynska N. The additive input-doubling method based on the SVR with nonlinear kernels: small data approach // *Symmetry*. 2021. Vol. 13, No. 4. P. 612. DOI: 10.3390/sym13040612. (Індексується у Scopus, квартиль Q2).

13. Izonin I., Tkachenko R., Berezhsky O., Krak I., Kováč M., Fedorchuk M. Improvement of the ANN-based prediction technology for extremely small biomedical data analysis // *Technologies*. 2024. Vol. 12, No. 7. P. 112. DOI: 10.3390/technologies12070112. (Індексується у Scopus, квартиль Q1).

14. Izonin I., Tkachenko R., Yendyk P., Pliss I., Bodyanskiy Y., Gregus M. Enhanced input-doubling method leveraging response surface linearization to improve classification accuracy in small medical data processing // *Computation*. 2024. Vol. 12, No. 10. P. 203. DOI: 10.3390/computation12100203. (Індексується у Scopus, квартиль Q2).

15. Izonin I., Tkachenko R., Hovdysh N., Berezhsky O., Yemets K., Tsmots I. Cascade-based input-doubling classifier for predicting survival in allogeneic bone marrow transplants: small data case // *Computation*. 2025. Vol. 13, No. 4. P. 80. DOI: 10.3390/computation13040080. (Індексується у Scopus, кuartиль Q2).
16. Izonin I., Tkachenko R., Bliakhar R., Kovac M., Bodyanskiy Y., Chala O. An improved ANN-based global-local approximation for small medical data analysis // *EAI Endorsed Transactions on Pervasive Health and Technology*. 2023. Vol. 9. DOI: 10.4108/eetpht.9.3320. (Індексується у Scopus, кuartиль Q2).
17. Izonin I., Tkachenko R., Yemets K., Havryliuk M. An interpretable ensemble structure with a non-iterative training algorithm to improve the predictive accuracy of healthcare data analysis // *Scientific Reports*. 2024. Vol. 14, No. 1. P. 12947. DOI: 10.1038/s41598-024-61776-y. (Індексується у Scopus, кuartиль Q1).
18. Izonin I., Muzyka R., Tkachenko R., Dronyuk I., Yemets K., Mitoulis S.-A. A method for reducing training time of ML-based cascade scheme for large-volume data analysis // *Sensors*. 2024. Vol. 24, No. 15. P. 4762. DOI: 10.3390/s24154762. (Індексується у Scopus, кuartиль Q1).
19. Izonin I., Tkachenko R., Krak I., Berezhsky O., Shevchuk I., Shandilya S. K. A cascade ensemble-learning model for the deployment at the edge: case on missing IoT data recovery in environmental monitoring systems // *Frontiers in Environmental Science*. 2023. Vol. 11. P. 1295526. DOI: 10.3389/fenvs.2023.1295526. (Індексується у Scopus, кuartиль Q2).
20. Izonin I., Tkachenko R., Gurbych O., Kovac M., Rutkowski L., Holoven R. A non-linear SVR-based cascade model for improving prediction accuracy of biomedical data analysis // *Mathematical Biosciences and Engineering*. 2023. Vol. 20, No. 7. P. 13398–13414. DOI: 10.3934/mbe.2023597. (Індексується у Scopus, кuartиль Q2).
21. Izonin I., Tkachenko R., Verhun V., Zub K. An approach towards missing data management using improved GRNN-SGTM ensemble method // *Engineering Science and Technology, an International Journal*. 2021. Vol. 24, No. 3. P. 749–759. DOI: 10.1016/j.jestch.2020.10.005. (Індексується у Scopus, кuartиль Q1).
22. Tkachenko R., Izonin I., Kryvinska N., Dronyuk I., Zub K. An approach towards increasing prediction accuracy for the recovery of missing IoT data based on the GRNN-SGTM ensemble // *Sensors*. 2020. Vol. 20, No. 9. Art. No. 9. DOI: 10.3390/s20092625. (Індексується у Scopus, кuartиль Q2).
23. Izonin I., Tkachenko R. Universal intraensemble method using nonlinear AI techniques for regression modeling of small medical data sets // *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*.

(Collective monograph). Amsterdam: Elsevier, 2022. P. 123–150. DOI: 10.1016/B978-0-323-85751-2.00002-5. *(Індексується у Scopus)*.

24. Ізонін І. В., Ткаченко Р. О., Зуб Х. В. Застосування бустингу в ансамблях нейронних мереж узагальненої регресії для підвищення точності розв'язання задач апроксимації // *На шляху до індустрії 4.0: інформаційні технології, моделювання, штучний інтелект, автоматизація*. (Колективна монографія). Одеса: Астропринт, 2021. С. 294–409.

25. Ізонін І. В., Ільчишин Б., Гаврилюк М., Бойчук О. Ансамбль однорідних методів машинного навчання з усередненням результату // *Інформаційні технології і автоматизація – 2023: матеріали XVI Міжнародної науково-практичної конференції*. Одеса, жовтень 2023. С. 339–340.

26. Ізонін І. В., Бейда Я. Р., Федорчук М. А., Машталір Д. Б., Підкостельний Р. Р. Інтелектуальний аналіз коротких вибірок даних: методи аугментації даних // *Економічний і соціальний розвиток України в XXI столітті: національна візія та виклики глобалізації: збірник тез доповідей XX Міжнародної науково-практичної конференції молодих вчених*. Тернопіль, травень 2023. С. 705–707.

27. Ізонін І. В., Бляхар Р. С., Ткаченко Р. О. Нейромережевий метод послідовної глобально-локальної апроксимації коротких наборів медичних даних // *Нейромережні технології та їх застосування: збірник наукових праць XXII Міжнародної наукової конференції НМТЗ-2023*. Краматорськ–Вінниця–Тернопіль, 2023. С. 87–89.

28. Izonin I., Tkachenko R., Greguš M. I-PNN: An Improved Probabilistic Neural Network for Binary Classification of Imbalanced Medical Data // In: Strauss C., Cuzzocrea A., Kotsis G., Tjoa A.M., Khalil I. (eds) *Database and Expert Systems Applications. DEXA 2022. Lecture Notes in Computer Science*, vol 13427. Cham: Springer, 2022. P. 147–157. DOI: 10.1007/978-3-031-12426-6\_12 *(Індексується у Scopus; квартиль Q3)*.

29. Izonin I., Tkachenko R., Ryvak L., Zub K., Rashkevych M., Pavliuk O. Addressing Medical Diagnostics Issues: Essential Aspects of the PNN-based Approach // *Proceedings of the 3rd International Conference on Informatics & Data-Driven Medicine (IDDM 2020)*. CEUR Workshop Proceedings, vol. 2753. Växjö, Sweden, 19–21 November 2020. P. 209–218. *(Індексується у Scopus)*.

30. Izonin I., Tkachenko R., Sydor M., Chretien S., Pliss I., Vovk T. Modification of Combined Unsupervised-Supervised Cascade Scheme for Small Biomedical Data Classification // *Proceedings of the 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*. Lviv, Ukraine: IEEE, September 2023. P. 107–112. DOI: 10.1109/ELIT61488.2023.10310830 *(Індексується у Scopus)*.

31. Izonin I., Tkachenko R., Shakhovska N., Ilchyshyn B., Gregus M., Strauss C. Towards Data Normalization Task for the Efficient Mining of Medical Data // Proceedings of the 12th International Conference on Advanced Computer Information Technologies (ACIT). Ruzomberok, Slovakia: IEEE, 2022. P. 480–484. DOI: 10.1109/ACIT54803.2022.9913112 (*Індексується у Scopus*).
32. Izonin I., Tepla T., Danylyuk D., Tkachenko R., Duriagina Z., Lemishka I. Towards an Intelligent Decision Making of Ti-based Powders Selection for Medical Manufacturing // Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA). Sakheer, Bahrain: IEEE, November 2020. P. 1–5. DOI: 10.1109/DASA51403.2020.9316866 (*Індексується у Scopus*).
33. Izonin I., Tkachenko R., Kryvinska N., Tkachenko P., Greguš M. Multiple Linear Regression Based on Coefficients Identification Using Non-iterative SGTm Neural-like Structure // In: Rojas I., Joya G., Catala A. (eds) Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science, vol 11506. Cham: Springer, 2019. P. 467–479. DOI: 10.1007/978-3-030-20521-8\_39 (*Індексується у Scopus, кuartиль Q2*).
34. Izonin I., Kryvinska N., Tkachenko R., Zub K., Vitynskyi P. An Extended-Input GRNN and its Application // Procedia Computer Science, vol. 160. Proceedings of the 10th EUSPN and 9th ICTH Conferences. Coimbra, Portugal: Elsevier, November 2019. P. 578–583. DOI: 10.1016/j.procs.2019.11.044 (*Індексується у Scopus*).
35. Tkachenko R., Izonin I. Model and Principles for the Implementation of Neural-Like Structures Based on Geometric Data Transformations // In: Hu Z., Petoukhov S., Dychka I., He M. (eds) Advances in Computer Science for Engineering and Education. ICCSEE 2019. Advances in Intelligent Systems and Computing, vol 754. Cham: Springer, 2019. P. 578–587. DOI: 10.1007/978-3-319-91008-6\_58 (*Індексується у Scopus, кuartиль Q3*).
36. Izonin I., Tkachenko R., Fedushko S., Koziy D., Zub K., Vovk O. RBF-Based Input Doubling Method for Small Medical Data Processing // In: Hu Z., Zhang Q., Petoukhov S., He M. (eds) Advances in Artificial Systems for Logistics Engineering. Lecture Notes on Data Engineering and Communications Technologies, vol 82. Cham: Springer, 2021. P. 23–31. DOI: 10.1007/978-3-030-80475-6\_3 (*Індексується у Scopus*).
37. Izonin I., Tkachenko R., Gregus M., Zub K., Lotoshynska N. Input Doubling Method based on SVR with RBF kernel in Clinical Practice: Focus on Small Data // Procedia Computer Science, vol. 184. Warsaw, Poland: Elsevier, March 2021. P. 606–613. DOI: 10.1016/j.procs.2021.03.075 (*Індексується у Scopus*).

38. Izonin I., Tkachenko R., Gregus M., Zub K., Tkachenko P. A GRNN-based Approach towards Prediction from Small Datasets in Medical Application // *Procedia Computer Science*, vol. 184. Warsaw, Poland: Elsevier, March 2021. P. 242–249. DOI: 10.1016/j.procs.2021.03.033 (*Індексується у Scopus*).
39. Izonin I., Tkachenko R., Pidkostelnyi R., Pavliuk O., Khavalko V., Batyuk A. Experimental Evaluation of the Effectiveness of ANN-based Numerical Data Augmentation Methods for Diagnostics Tasks // *Proceedings of the 4th International Conference on Informatics & Data-Driven Medicine (IDDM 2021)*. CEUR Workshop Proceedings, vol. 3038. Valencia, Spain, November 2021. P. 223–232. (*Індексується у Scopus*).
40. Izonin I., Tkachenko R., Horbal N., Greguš M., Verhun V., Tolstyak Y. An Approach Toward Numerical Data Augmentation and Regression Modeling Using Polynomial-Kernel-Based SVR // In: Saraswat M., Roy S., Chowdhury C., Gandomi A.H. (eds) *Proceedings of International Conference on Data Science and Applications. Lecture Notes in Networks and Systems*, vol 288. Singapore: Springer, 2022. P. 771–781. DOI: 10.1007/978-981-16-5120-5\_58 (*Індексується у Scopus, квартиль Q4*).
41. Izonin I., Tkachenko R., Yemets K., Gregus M., Tomashy Y., Pliss I. An Approach Towards Reducing Training Time of the Input Doubling Method via Clustering for Middle-Sized Data Analysis // *Procedia Computer Science*, vol. 241. Huntington, WV, USA: Elsevier, 2024. P. 32–39. DOI: 10.1016/j.procs.2024.08.007 (*Індексується у Scopus*).
42. Izonin I., Tkachenko R., Trush B., Nahirnyi S. Homogeneous ensemble models based on ANNs with noniterative training: classification and regression tasks // *Book of Abstracts of Workshops RECI 2022 and ACeSYRI 2022*. Delft, Netherlands, 15 November 2022. P. 27. URL: <https://ki.fri.uniza.sk/RECI2022/Abstracts.pdf>.
43. Izonin I., Tkachenko R., Vitynskyi P., Zub K., Tkachenko P., Dronyuk I. Stacking-based GRNN-SGTM Ensemble Model for Prediction Tasks // *Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA)*. Sakheer, Bahrain: IEEE, November 2020. P. 326–330. DOI: 10.1109/DASA51403.2020.9317124 (*Індексується у Scopus*).
44. Tkachenko R., Izonin I., Greguš M., Tkachenko P., Dronyuk I. Committee of the SGTM Neural-Like Structures with Extended Inputs for Predictive Analytics in Insurance // In: Younas M., Awan I., Benbernou S. (eds) *Big Data Innovations and Applications. Communications in Computer and Information Science*, vol 1054. Cham: Springer, 2019. P. 121–132. DOI: 10.1007/978-3-030-27355-2\_9 (*Індексується у Scopus, квартиль Q3*).

45. Tkachenko R., Tkachenko P., Izonin I., Vitynskyi P., Kryvinska N., Tsymbal Y. Committee of the Combined RBF-SGTM Neural-Like Structures for Prediction Tasks // In: Awan I., Younas M., Ünal P., Aleksy M. (eds) Mobile Web and Intelligent Information Systems. Lecture Notes in Computer Science, vol 11673. Cham: Springer, 2019. P. 267–277. DOI: 10.1007/978-3-030-27192-3\_21 (*Індексується у Scopus, квартиль Q2*).
46. Izonin I., Tkachenko R., Kryvinska N., Gregus M., Tkachenko P., Vitynskyi P. Committee of SGTM Neural-Like Structures with RBF kernel for Insurance Cost Prediction Task // 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON). Lviv, Ukraine: IEEE, July 2019. P. 1037–1040. DOI: 10.1109/UKRCON.2019.8879905 (*Індексується у Scopus*).
47. Tkachenko R., Izonin I., Kryvinska N., Chopyak V., Lotoshynska N., Danylyuk D. Piecewise-linear Approach for Medical Insurance Costs Prediction using SGTM Neural-Like Structure // Proceedings of the 1st International Workshop on Informatics & Data-Driven Medicine (IDDM 2018). CEUR Workshop Proceedings, vol. 2255. Lviv, Ukraine, November 2018. P. 170–179. (*Індексується у Scopus*).
48. Izonin I., Tkachenko R., Shcherbii O., Berezhsky O., Krak I., Oliinyk M. An Approximation Cascade Scheme via Rational Fractions for Biomedical Data Analysis // Proceedings of the 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT). Lviv, Ukraine: IEEE, October 2023. P. 1–4. DOI: 10.1109/CSIT61576.2023.10324122 (*Індексується у Scopus*).
49. Izonin I., Tkachenko R., Shcherbii O., Muzyka R., Faramarzi A., Mitoulis S.A. Enhanced Cascade Schemes for Advancing Machine Learning-Based Prediction of Heating and Cooling Loads in Residential Buildings // Proceedings of the 2024 Mediterranean Smart Cities Conference (MSCC). Martil–Tetuan, Morocco: IEEE, May 2024. P. 1–5. DOI: 10.1109/MSCC62288.2024.10697024 (*Індексується у Scopus*).
50. Izonin I., Tkachenko R. An approach towards the response surface linearization via ANN-based cascade scheme for regression modeling in Healthcare // Procedia Computer Science, vol. 198. Leuven, Belgium: Elsevier, 2022. P. 724–729. DOI: 10.1016/j.procs.2021.12.313 (*Індексується у Scopus*).
51. Ізонін І. В., Ткаченко Р. О., Томаши Є. О., Гентош Л. І. Бібліотека методів інтелектуального аналізу малих даних на основі декартового добутку впорядкованих множин входів: комп'ютерна програма. Свідоцтво про реєстрацію авторського права № 127079. Зареєстровано 4 червня 2024 р. – Заявник: Ізонін І. В. – Номер заявки: с202403578; дата подання: 23 квітня 2024 р.

52. Вітинський П. Б., Ткаченко Р. О., Ізонін І. В. Комбінована нейроподібна система для задач аналітики в умовах коротких вибірок даних: комп'ютерна програма. Свідоцтво про реєстрацію авторського права № 93094. Зареєстровано 16 жовтня 2019 р. – Заявник: Вітинський П. Б., Ткаченко Р. О., Ізонін І. В. – Номер заявки: 94679; дата подання: 25 вересня 2019 р.

53. Ізонін І. В., Нагірний С. Б., Ткаченко Р. О. Гомогенні стекінгові ансамблі штучних нейронних мереж без навчання (Multi-GRNN): комп'ютерна програма. Свідоцтво про реєстрацію авторського права № 114033. Зареєстровано 9 серпня 2022 р. – Заявник: Ізонін І. В. – Номер заявки: с202203246; дата подання: 25 липня 2022 р.

## АНОТАЦІЯ

*Ізонін І.В.* Ансамблі моделей неітеративного машинного навчання для аналізу біомедичних даних малих обсягів. – Кваліфікаційна наукова праця на правах рукопису.

Дисертаційна робота на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – Системи та засоби штучного інтелекту. – Національний університет “Львівська політехніка” Міністерства освіти і науки України, Львів, 2025.

У дисертаційній роботі вирішено важливу теоретичну проблему розвитку основ теорії синтезу нейроподібних ансамблевих структур з неітеративним машинним навчанням для підвищення точності розв’язання задач регресії та класифікації на вибірках біомедичних даних малих обсягів.

Проведено аналіз методів інтелектуального опрацювання структурованих біомедичних даних малого обсягу та обґрунтовано доцільність синтезу нейроподібних ансамблевих структур на основі неітеративного машинного навчання. Розвинуто підхід до використання кусково-лінійної апроксимації в задачах аналізу багатовимірних поверхонь відгуку шляхом ізоморфного подання кожної ознаки множиною локальних компонент. Це дало змогу здійснити нелінійне розширення вхідного простору і знизити похибку апроксимації лінійною неітеративною структурою Модель Послідовних Геометричних Перетворень (МПГП) на 58 % за показником RMSE. Удосконалено метод формування комбінованої поверхні відгуку із застосуванням кластерного аналізу з мінімальним зростанням вимірності простору ознак, що забезпечило збереження узагальнюючої здатності та підвищення точності моделей. Розроблено метод підвищення вимірності простору незалежних ознак вибірки шляхом використання вихідних сигналів шару підсумовування ймовірнісної нейронної мережі, який підвищив точність класифікації на 43 % і 9 % за F1-мірою відповідно до базових ШНМ. Розроблено нову методологію аугментації екстремально коротких вибірок біомедичних даних на основі декартового квадрата множини вхідних

компонент векторів, що дозволило застосовувати різнотипні нелінійні моделі, зокрема, МППП з РБФ-розширенням входів для реалізації процедури навчання. Завдяки цьому забезпечено підвищення точності розв'язання задач регресії (зниження RMSE до 59 %). Розроблено методи синтезу ансамблів неітеративних лінійних і нелінійних нейроподібних структур з використанням кусково-лінійної апроксимації, РБФ-розширення входів, симетричних випадкових зміщень з подальшим агрегуванням вихідних сигналів усіх членів ансамблю мета-моделлю, що дозволило підвищити точність результатів розв'язання задач регресії та класифікації. Розроблено метод часткового усунення систематичної похибки шляхом використання ансамблю з двох ШНМ узагальненої регресії, що забезпечив зменшення RMSE на 4–8 %. Удосконалено метод глобально-локальної апроксимації, який підвищив точність на 13 %. Розроблено прямий та зворотний методи компенсації похибки моделі регресії на основі застосування раціональних дробів, що забезпечили зниження MSE на 9–11 %.

Усі вищезазначені методи реалізовано в авторських програмних засобах, які захищено свідоцтвами про реєстрацію авторського права. Проведено експериментальну оцінку їх ефективності на модельних і реальних біомедичних даних. Достовірність отриманих результатів підтверджується результатами порівняння з існуючими методами машинного навчання та шляхом апробації в низці прикладних задач біоматеріалознавства, ревматології, травматології, стоматології та трансплантології. Впровадження здійснювалося у співпраці з медичними установами, науковими закладами та в межах міжнародних грантових проєктів, що підтверджено відповідними актами впровадження та використання.

Ключові слова: біомедичні дані малих обсягів, штучний інтелект, ансамблеве навчання, штучна нейронна мережа, неітеративне навчання, аугментація даних, нелінійне розширення простору вхідних даних, каскадні методи, компенсація похибок моделі регресії, глобально-локальна апроксимація, раціональні дроби.

## ABSTRACT

Izonin I.V. Ensembles of non-iterative machine learning models for small biomedical data analysis. – On the rights of the manuscript.

Dissertation for the degree of Doctor of Technical Sciences in specialty 05.13.23 – Systems and Means of Artificial Intelligence. – Lviv Polytechnic National University of the Ministry of Education and Science of Ukraine, Lviv, 2025.

This dissertation addresses a significant theoretical problem of advancing the foundations of synthesizing neural-like ensemble structures based on non-iterative machine learning algorithm to improve the accuracy of solving regression and classification tasks on small-volume biomedical datasets. An analysis of existing methods for the intelligent processing of structured small-scale biomedical data is

conducted, and the need for constructing neural-like ensemble structures using non-iterative machine learning algorithm is substantiated.

A novel approach is proposed for utilizing piecewise-linear approximation in the analysis of multidimensional response surfaces by means of isomorphic representation of each input feature as a set of local components. This allowed for nonlinear expansion of the input space and reduction of approximation error by 58% in RMSE using a linear non-iterative SGTm neural-like structure. A method for forming a combined response surface based on cluster analysis is improved, where the original feature space is extended solely by the cluster center's target value. This minimized the increase in dimensionality while improving generalization and prediction accuracy. A method is developed for expanding the space of independent features in the dataset through the use of output signals from the summation layer of a Probabilistic Neural Network, which led to a 43% and 9% increase in classification accuracy (according to F1-score) compared to baseline artificial neural networks.

A new methodology is developed for augmenting extremely small biomedical datasets using Cartesian squaring of the input vector components, which enabled the application of various nonlinear machine learning models, particularly SGTm neural-like structures with RBF-based input expansion. This approach improved regression accuracy (RMSE reduced to 59%). Methods for synthesizing ensembles of non-iterative linear and nonlinear neural-like models were proposed using piecewise-linear approximation, RBF-based input expansion, and symmetric random shifts, followed by aggregation of their outputs via a meta-model. A method for partial correction of systematic error in General Regression Neural Networks is developed using an ensemble of two such networks, achieving a 4–8% RMSE reduction. The method of global-local approximation was enhanced, yielding a 13% increase in accuracy. Additionally, direct and inverse error compensation methods for regression models were proposed based on rational function transformations, resulting in 9–11% reduction in MSE.

All of the above-mentioned methods have been implemented in proprietary software tools protected by certificates of copyright registration. Their effectiveness has been experimentally validated using both synthetic and real biomedical data. The reliability of the obtained results is supported by comparative evaluation against existing machine learning methods and by successful approbation in a number of applied tasks in biomaterials science, rheumatology, traumatology, dentistry, and transplantology. The implementation was carried out in collaboration with medical institutions, academic organizations, and within the framework of international grant projects, which is confirmed by relevant implementation and utilization acts.

*Keywords:* small biomedical data, artificial intelligence, ensemble learning, artificial neural network, non-iterative learning, data augmentation, nonlinear feature space expansion, cascade methods, regression error compensation, global-local approximation, rational functions.