

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0826U001977

Особливі позначки: відкрита

Дата реєстрації: 27-05-2026

Статус: Захищена

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Гаврилюк Мирослав Андрійович

2. Myroslav A. Havryliuk

Кваліфікація:

Ідентифікатор ORCID ID: 0000-0001-5259-7564

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 03-07-2026

Спеціальність за освітою: 122 Комп'ютерні науки

Місце роботи здобувача: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 13880

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.54.06

Тема дисертації:

1. Методи та засоби інтелектуального аналізу коротких наборів медичних даних
2. Methods and means for intelligent analysis of short medical datasets

Реферат/анотація:

1. У дисертаційній роботі розв'язано актуальну наукову задачу підвищення здатності до узагальнення моделей регресії та класифікації, побудованих на основі медичних даних малих обсягів, шляхом розроблення й удосконалення методів інтелектуального аналізу даних та розвитку ансамблів таких моделей, що забезпечують врахування особливостей коротких, високорозмірних та незбалансованих вибірок. Розроблені, удосконалені та розвинуті методи орієнтовані на поєднання аугментації та ансамблевого навчання для коротких наборів даних із різних галузей медицини. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку літературних джерел та додатків. У першому розділі виконано аналітичний огляд моделей і методів розв'язання задач регресії та класифікації у випадку коротких наборів медичних даних. Проаналізовано причини нестачі доступних для аналізу даних у медичній сфері. Описано особливості коротких наборів даних та проблеми, які виникають під час їх інтелектуального аналізу. Наведено визначення поняття коротка вибірка медичних даних у межах дисертаційного дослідження. Формально описано задачі регресії та класифікації. Проведено огляд існуючих моделей для аналізу коротких

наборів даних. Розглянуто методи, засновані на принципах декартового квадрата множини вхідних компонент векторів для розв'язання задач регресії та класифікації. Проаналізовано причини підвищення ефективності інтелектуального аналізу за їх допомогою у випадку коротких вибірок даних. У другому розділі розроблено та удосконалено методи інтелектуального аналізу коротких високорозмірних медичних вибірок для розв'язання задач регресії, бінарної та мультикласової класифікації шляхом модифікації процедури аугментації на основі декартового квадрата множини вхідних компонент векторів. Для усунення зазначених обмежень удосконалено метод декартового квадрата шляхом переходу від конкатенації векторів до покомпонентного обчислення різниць між усіма парами вхідних екземплярів із додатковим використанням каскадного принципу формування ознак. Такий підхід дозволяє зменшити розмірність аугментованої вибірки та підвищити інформативність представлення даних. Теоретично обґрунтовано, що удосконалений метод забезпечує підвищення здатності до узагальнення нелінійних моделей регресії та бінарної класифікації при аналізі коротких високовимірних вибірок даних. Крім того, розроблено метод розв'язання задач мультикласової класифікації з адитивною компенсацією вихідного сигналу. Він передбачає формування штучного цільового атрибуту шляхом додавання зваженої суми вхідних ознак до значення мітки класу з подальшим застосуванням попарного представлення даних. Такий підхід забезпечує коректну інтерпретацію цільового сигналу та розширює область застосування підходів на основі декартового квадрата на багатокласові задачі. У третьому розділі розвинуто методи інтелектуального аналізу коротких незбалансованих вибірок медичних даних на основі ансамблів моделей декартового квадрата для розв'язання задач регресії та бінарної класифікації. Зокрема, подальшого розвитку набуло застосування ансамблевого підходу типу бегінг у задачах регресії, у межах якого для кожної випадкової підвибірки синтезується окремий базовий член ансамблю із використанням удосконаленого методу декартового квадрата множини вхідних компонент векторів. Крім того, розвинуто ансамбль збалансованих підвибірок даних для аналізу даних в умовах дисбалансу класів у задачі бінарної класифікації. Тут кожен базовий член ансамблю навчається на власній збалансованій підвибірці з вирівнюванням обсягів класів до рівня міноритарного, а результати першого рівня узгоджуються із використанням метамоделі. Розвинутий метод зменшує зміщення оцінок апостеріорних ймовірностей, спричинене домінуванням класу більшості, та забезпечує підвищення здатності до узагальнення моделей при аналізі коротких наборів медичних даних. У четвертому розділі розроблено програмно-алгоритмічний комплекс, за допомогою якого можна виконувати інтелектуальний аналіз коротких вибірок медичних даних. Апробацію виконано на прикладі низки практичних задач зі сфери медицини. Результати продемонстрували підвищення здатності до узагальнення обох алгоритмічних реалізацій розробленої моделі порівняно з базовими. Окрім цього, розвинені ансамблі моделей на основі декартового квадрата множини вхідних компонент векторів верифіковано та апробовано при розв'язанні задач регресії та класифікації. Результати порівняння з існуючими ансамблями продемонстрували суттєво кращу здатність до узагальнення розвинених ансамблів при розв'язанні практичних задач із галузей ревматології, ендокринології травматології та трансплантології.

2. The dissertation solves the urgent scientific problem of increasing the ability to generalize regression and classification models built on the basis of small-scale medical data by developing and improving data mining methods and developing ensembles of such models that take into account the features of short, high-dimensional and unbalanced samples. The developed, improved and advanced methods are focused on combining augmentation techniques and ensemble learning for small data sets from various fields of medicine. The dissertation consists of an introduction, four sections, conclusions, a list of literature sources and appendices. The first section provides an analytical review of models and methods for solving regression and classification problems in the case of short sets of medical data. The reasons for the lack of data available for analysis in the medical field are analyzed. The features of short data samples and the problems that arise during their intellectual analysis are described. The definition of the concept of "short sample of medical data" within the framework of the dissertation research is given. The regression and classification tasks are formally described. The existing models for the analysis of short data sets are reviewed. Methods based on the principles of the Cartesian square of the set of input vector components for solving regression and classification problems are considered. They are focused on

the development and improvement of methods based on the Cartesian square of the set of input vector components. In the second section, methods for intelligent analysis of short high-dimensional medical samples are improved and developed for solving regression, binary and multiclass classification problems by modifying the augmentation procedure based on the Cartesian square of the set of input vector components. To eliminate these limitations, the Cartesian square method is improved by switching from vector concatenation to component-by-component calculation of differences between all pairs of input instances with the additional use of the cascade principle of feature formation. This approach allows you to reduce the dimension of the augmented sample and increase the informativeness of the data presentation. It is theoretically substantiated that the improved method provides an increase in the ability to generalize nonlinear regression models and binary classification when analyzing short highdimensional data samples. In addition, a method for solving multiclass classification problems with additive compensation of the output signal has been developed. It involves the formation of an artificial target attribute by adding a weighted sum of input features to the class label value with the subsequent application of pairwise data representation. This approach ensures the correct interpretation of the target signal and expands the scope of application of approaches based on the Cartesian square to multiclass problems. In the third section, methods for intelligent analysis of short unbalanced samples of medical data based on ensembles of Cartesian square models for solving regression and binary classification problems are developed. The use of an ensemble approach of the bagging type in regression problems has been further developed, within which a separate basic member of the ensemble is synthesized for each random subsample using the improved Cartesian square method of the set of input vector components. In addition, a balanced class ensemble is developed for data analysis under class imbalance in a binary classification problem, in which each base member is trained on its own balanced subsample with class sizes equalized to the minority class, and the first-level results are reconciled using metaaggregation. The developed method reduces the bias in posterior probability estimates caused by the dominance of the majority class and improves the generalizability of models when analyzing small-volume medical data samples. In the fourth section, a software and algorithmic complex is developed, with the help of which it is possible to perform intelligent analysis of short samples of medical data. The test is carried out on the example of a number of practical problems in the field of medicine. The results demonstrated an increase in the ability to generalize both algorithmic implementations of the developed model compared to the basic ones. In addition, the developed ensembles of models based on the Cartesian square of the set of input vector components were verified and tested for solving regression and classification problems. The results of comparison with existing ensembles demonstrated a significantly better ability to generalize the developed ensembles for solving practical problems in the fields of rheumatology, endocrinology, traumatology and transplantology.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- 1. Ramakrishna M., Izonin I., Havryliuk M., Venkatesan V., Bhat C. Homogeneous adaboost ensemble machine learning algorithms with reduced entropy on balanced data. Entropy. 2023. Vol. 25, iss. 2. 245. doi: 10.3390/e25020245
- 2. Izonin I., Tkachenko R., Havryliuk M., Venkatesan V., Ramakrishna M. Efficient data preprocessing with ensemble machine learning technique for the early detection of chronic kidney disease. Applied Sciences (Basel). 2023. Vol.13, iss. 5. 2885. doi: 10.3390/app13052885
- 3. Izonin I., Stupnytskyi M., Tkachenko R., Havryliuk M., Biletskyi O., Melnyk G. Mortality risk prediction for multiple trauma patients admitted to the hospital via machine learning algorithms. Lecture Notes on Data

Engineering and Communications Technologies. 2025. Vol. 242 : Advances in computer science for engineering and education VII (ICCSEEA 2024). Vol. 1. P. 218–228. doi: 10.1007/978-3-031-84228-3_18

- 4. Izonin I., Tkachenko R., Havryliuk M., Yendyk P., Tolstyak Y., Gregus M. An adaptation of the input doubling method for solving classification tasks in case of small data processing. *Procedia Computer Science*. 2024. Vol. 241 : 19th Intern. conf. on future networks and communications, 21th Intern. conf. on mobile systems and pervasive computing, 14th Intern. conf. on sustainable energy information technology: proc., Huntington, USA, 5–7 Aug. 2024. P. 171–178. doi:10.1016/j.procs.2024.08.024
- 5. Havryliuk M., Hovdysh N., Tolstyak Y., Chopyak V., Kustra N. Investigation of PNN optimization methods to improve classification performance in transplantation medicine. *CEUR Workshop Proceedings*. 2023. Vol. 3609 : 6th Intern. conf. on informatics and data-driven medicine, IDDM 2023, Bratislava, Slovakia, Nov. 17–19, 2023. P. 338–345. <https://ceur-ws.org/Vol-3609/short9.pdf>
- 6. Havryliuk M. Enhanced two-step augmentation method for analyzing small datasets in medical applications. *Комп'ютерні системи та інформаційні технології*. 2025. № 1. С. 156–162. doi: 10.31891/csit-2025-1-18
- 7. Havryliuk M. Method of intellectual analysis of short high-dimensional samples based on bagging ensemble with data augmentation. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2025. № 3. С. 144–151. doi: 10.31891/2219-9365-2025-83-19
- 8. Гаврилюк М. А. Стекінговий підхід для подолання викликів малих незбалансованих медичних даних. *Наука і техніка сьогодні*. 2025. № 8 (49). С. 1201–1213. doi: 10.52058/2786-6025-2025-8(49)-1201-1213
- 9. Гаврилюк М. А., Говдиш Н. О. Дослідження ефективності використання PNN для розв'язання задач медичної діагностики в умовах аналізу малих даних високої розмірності. *Вісник Хмельницького національного університету. Серія: Технічні науки*. 2024. № 1 (331). С. 248–251. doi: 10.31891/2307-5732-2024-331-37
- 10. Гаврилюк М. А., Ізонін І. В., Ткаченко Р. О. Метод розв'язання задач мультикласової класифікації на основі декартового квадрату впорядкованих множин входів з адитивною компенсацією вихідного сигналу. *Наука і техніка сьогодні*. 2026. № 1 (55). С. 2072–2084. doi: 10.52058/2786-6025-2026-1(55)-2072-2084

Наукова (науково-технічна) продукція: методи, теорії, гіпотези

Соціально-економічна спрямованість: забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПВ:

Комп'ютерні програми

Комп'ютерна програма "Бібліотека методів інтелектуального аналізу малих даних на основі принципів осьової симетрії поверхні відгуку". Свідectvo про реєстрацію авторського права на твір №1227078 від 04.06.2024р.

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Ізонін Іван Вікторович

2. Ivan V. Izonin

Кваліфікація: д. т. н., доц., 05.13.23

Ідентифікатор ORCID ID: 0000-0002-9761-0096

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Манзюк Едуард Андрійович

2. Eduard A. Manziuk

Кваліфікація: д. т. н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0002-7310-2126

Додаткова інформація:

Повне найменування юридичної особи: Хмельницький національний університет

Код за ЄДРПОУ: 02071234

Місцезнаходження: вул. Інститутська, Хмельницький, Хмельницький р-н., 29016, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Шафроненко Аліна Юріївна

2. Alina Shafronenko

Кваліфікація: д. т. н., професор, 05.13.23

Ідентифікатор ORCID ID: 0000-0002-8040-0279

Додаткова інформація:

Повне найменування юридичної особи: Харківський національний університет радіоелектроніки

Код за ЄДРПОУ: 02071197

Місцезнаходження: проспект Науки, Харків, Харківський р-н., 61166, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Мельникова Наталія Іванівна

2. Nataliia I. Melnykova

Кваліфікація: д. т. н., доц., 05.13.23

Ідентифікатор ORCID ID: 0000-0002-2114-3436

Додаткова інформація: Web of Science Researcher ID: E-1561-2019; Scopus Author ID: 55312617400;
<https://scholar.google.com/citations?user=IYuuQSsAAAAJ&hl=uk>

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Ляковська Соломія Євгенівна

2. Solomia E. Liaskovska

Кваліфікація: к. т. н., доц., 05.01.01

Ідентифікатор ORCID ID: 0000-0002-0822-0951

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові
голови ради

Яковина Віталій Степанович

Власне Прізвище Ім'я По-батькові
головуючого на засіданні

Яковина Віталій Степанович

**Відповідальний за підготовку
облікових документів**

Реєстратор

Віктор Михайлович Хавалко

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є
відповідальним за реєстрацію наукової
діяльності**



Юрченко Тетяна Анатоліївна