

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

БУЛАХ ВІТАЛІЙ АНАТОЛІЙОВИЧ

УДК 004.9; 519.2



**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСИФІКАЦІЇ
ВПОРЯДКОВАНИХ МАСИВІВ ДАНИХ ІЗ ФРАКТАЛЬНИМИ
ВЛАСТИВОСТЯМИ МЕТОДАМИ МАШИННОГО НАВЧАННЯ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2021

Дисертацією є рукопис.

Робота виконана у Харківському національному університеті радіоелектроніки Міністерства освіти і науки України.

Науковий керівник доктор технічних наук, професор
Кіріченко Людмила Олегівна,
Харківський національний університет
радіоелектроніки,
професор кафедри прикладної математики.

Офіційні опоненти: доктор технічних наук, професор
Жолткевич Григорій Миколайович,
Харківський національний університет
імені В.Н. Каразіна,
декан факультету математики та інформатики;

доктор технічних наук, професор
Шаронова Наталія Валеріївна,
Національний технічний університет
«Харківський політехнічний інститут»,
завідувачка кафедри інтелектуальних
комп'ютерних систем.

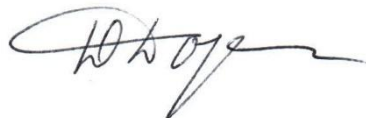
Захист відбудеться «29» квітня 2021 р. о 16 годині на засіданні спеціалізованої вченої ради Д 64.050.07 в Національному технічному університеті «Харківський політехнічний інститут» за адресою: 61002, м. Харків, вул. Кирпичова, 2.

З дисертацією можна ознайомитися у бібліотеці Національного технічного університету «Харківський політехнічний інститут» за адресою: 61002, м. Харків, вул. Кирпичова, 2.

Автореферат розісланий «26» березня 2021 р.

Вчений секретар

спеціалізованої вченої ради



Юрій ДОРОФЄЄВ

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Із відкриттям у 1977 році Б. Мандельбротом фрактальної геометрії у природі, дослідники почали виявляти фрактальні властивості не лише в геометричних та природних системах, але й в економіці, фізиці, медицині та техніці. Теперішнім часом результати фрактального аналізу широко використовуються в технічних, фізичних, біологічних, інформаційних системах. Треба відзначити, що методи фрактального аналізу даних є досить складними, оцінювання фрактальних характеристик за малим обсягом даних призводить до широкого довірчого інтервалу, що ускладнює інтерпретацію результатів і прийняття рішень.

У багатьох випадках виникають задачі розпізнавання і класифікації фрактальних упорядкованих масивів даних (УМД). В останні роки зростає інтерес до методів машинного навчання для аналізу та класифікації фрактальних даних. Методи машинного навчання дозволяють класифікувати дані, базуючись лише на їх статистичних характеристиках, або не вимагають високої точності оцінки фрактальних характеристик. Використання методів машинного навчання дозволяє компенсувати похибки оцінювання фрактальних та інших статистичних характеристик при побудові класифікатора.

Однак, на сьогоднішній день не розроблено інформаційної технології (ІТ), яка призначена для задач класифікації упорядкованих масивів даних з фрактальними властивостями. Така ІТ повинна задовольняти наступним вимогам:

- ІТ має збирати вхідні дані для аналізу в онлайн-режимі та зберігати їх до бази даних;
- в умовах недостатньої кількості даних ІТ має використовувати методи моделювання для генерації даних із подібними фрактальними властивостями до аналізованих процесів;
- використовуючи методи машинного навчання, згенеровані та отримані реальні дані, ІТ повинна навчитися оптимально виявляти (класифікувати) різні стани досліджуваного об'єкта;
- використовуючи навчену модель, класифікувати вхідні дані, інформувати зовнішню систему та оператора ІТ під час виявлення особливостей вхідних даних;
- зберігати до бази даних всі результати роботи ІТ;
- зі збільшенням обсягів реальних даних проводити перенавчання моделі класифікації.

Таким чином актуальність роботи полягає в розробці сукупності ефективних взаємопов'язаних методів та побудові інформаційної технології класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання.

Зв'язок роботи з науковими програмами, планами, темами.

Результати дисертаційної роботи використані в ході виконання науково-дослідницьких робіт Харківського національного університету

радіоелектроніки (ХНУРЕ) у межах договору за замовленням Міністерства освіти і науки України (ДР № 01190U001405) «Автоматизована оптична інформаційно-вимірювальна система для полігонних випробувань керованих та некерованих ракет, артилерійських і реактивних снарядів», в яких здобувач був виконавцем окремих етапів.

Мета та задачі дослідження. Метою дисертаційної роботи є підвищення точності класифікації упорядкованих масивів даних із фрактальними властивостями шляхом створення інформаційної технології класифікації методами машинного навчання.

Для досягнення зазначеної мети поставлені наступні задачі:

- провести аналіз існуючих методів класифікації упорядкованих масивів даних методами машинного навчання;
- розробити методи генерації упорядкованих масивів даних з фрактальними властивостями різних типів для побудови навчальних вибірок під час виконання навчання моделі класифікації з учителем;
- виконати експериментальні дослідження розроблених моделей даних та методів класифікації для визначення їх ефективності;
- запропонувати інформаційну технологію класифікації фрактальних упорядкованих масивів даних

Об'єкт дослідження – процес класифікації об'єктів методами машинного навчання.

Предмет дослідження – методи та моделі класифікації фрактальних упорядкованих масивів даних.

Методи дослідження. При вирішенні поставлених задач у роботі використовувався математичний апарат методів машинного навчання; методи математичної статистики – для статистичного аналізу упорядкованих масивів даних, аналізу точності класифікації методами машинного навчання; методи теорії хаоса та фрактального аналізу – для обчислення фрактальних та рекурентних характеристик.

Наукова новизна отриманих результатів полягає у тому, що вперше встановлено наступне:

- розроблено інформаційну технологію класифікації фрактальних упорядкованих масивів даних, яка заснована на поєднанні методів дерев рішень та нейронних мереж, враховує самоподібні та мультифрактальні властивості упорядкованих даних, що дозволяє збільшити точність класифікації;
- запропоновано метод оцінювання показника Херста за значеннями часового ряду, який засновано на застосуванні класифікаційних методів машинного навчання, що дозволяє суттєво зменшити довірчий інтервал оцінки;
- застосовані рекурентні характеристики часових рядів, як ознаки для класифікації часових рядів на основі їхніх фрактальних властивостей методами машинного навчання, дозволили збільшити точність класифікації;

Подальший розвиток отримали:

- метод генерації фрактальних реалізацій на основі стохастичного мультиплікативного біноміального каскаду, який на відміну від існуючих

використовує аналітичне визначення параметрів несиметричного бета-розподілу, що дозволяє генерувати упорядковані масиви даних із заданими мультифрактальними властивостями;

– методи бінарної класифікації часових рядів з фрактальними властивостями, що засновані на використанні дерев прийняття рішень та нейронних мереж, які на відміну від існуючих використовують як ознаки фрактальні та рекурентні характеристики, що дозволяє збільшити точність класифікації.

Практичне значення отриманих результатів полягає у тому, що в результаті виконаного дисертаційного дослідження створено комплекс методів, алгоритмів і програм, які дозволяють вирішити завдання класифікації фрактальних упорядкованих масивів даних. Усі теоретичні розробки дисертації доведені автором до конкретних методів і покладені в основу функціонування інформаційної технології класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання.

Запропонований метод оцінювання показника Херста, який програмно реалізований, дозволяє покращити точність його оцінювання за упорядкованими даними, а саме: зменшує довірчий інтервал оцінки в декілька разів.

Результати дисертаційного дослідження впроваджені на підприємстві ПрАТ «Фарлеп-Інвест» (м. Харків) для зменшення негативного впливу кібератак на інформаційні системи підприємства, на підприємстві КБ «ПриватБанк» (м. Харків) для виявлення вторгнень в інформаційні системи підприємства та у навчальному процесі в Харківському національному університеті радіоелектроніки при викладанні дисципліни «Artificial Intelligence (Штучний інтелект)».

Створена інформаційна технологія може бути використана для класифікації різноманітних упорядкованих даних і сигналів з фрактальними властивостями, зокрема – детектування кібереатак в інфокомунікаційних трафіках, уточнення діагнозу за даними записів електроенцефалограми та електрокардіографії, класифікації сейсмічних подій за сейсмограмами тощо.

Особистий внесок здобувача. Положення та результати, винесені на захист дисертаційної роботи, отримані здобувачем особисто. Серед них: програмно реалізовано методи генерації упорядкованих даних з фрактальними властивостями різних типів: фрактальний броунівський рух, експоненціальне перетворення фрактального гаусівського шуму, стохастичний біноміальний мультиплікативний каскад та мультифрактальний броунівський рух; удосконалено метод генерації мультифрактальних реалізацій на основі стохастичного мультиплікативного біноміального каскаду; запропоновано використовувати рекурентні характеристики УМД як ознаки для класифікації методами машинного навчання; запропоновано метод оцінювання показника Херста за значеннями УМД, який засновано на застосуванні класифікаційних методів машинного навчання; розроблено та програмно реалізовано

інформаційну технологію класифікації фрактальних упорядкованих масивів даних, яка заснована на поєднанні методів дерев рішень та нейронних мереж, а також враховує самоподібні та мультифрактальні властивості упорядкованих даних.

Апробація результатів дисертації. Основні положення та результати роботи доповідались на: 4th International scientific-practical conference «Problems of infocommunications. Science and technology» (Kharkiv, 2017); 6-й Міжнародній науково-практичній конференції «Практичне застосування нелінійних динамічних систем в інфокомунікаціях» (Чернівці, 2017); 9th IEEE International conference «Dependable systems, services and technologies» (Kyiv, 2018); 14th international conference on ICT in education, research and industrial applications «Integration, harmonization and knowledge transfer» (Kyiv, 2018); IEEE Second international conference «Data stream mining & processing» (Lviv, 2018); International scientific-practical conference «Problems of infocommunications. Science and technology» (Kharkiv, 2018); 7-й Міжнародній науково-практичній конференції «Фізико-технічні проблеми передавання, обробки та зберігання інформації в інфокомунікаційних системах» (Чернівці, 2018); Міжнародній науково-технічній конференції «Інформаційні технології в металургії та машинобудуванні» (Дніпро, 2018); 10th IEEE international conference «Intelligent data acquisition and advanced computing systems: technology and applications» (Metz, 2019); IEEE 8th international conference «Advanced optoelectronics and lasers» (Sozopol, 2019); 5-й Міжнародній науково-практичній конференції «Інтелектуальні рішення» (Ужгород, 2019); Third international workshop «Computer modeling and intelligent systems» (Zaporizhzhia, 2020).

Публікації. Матеріали дисертації опубліковано в 24 наукових публікаціях, з них: 3 у наукових спеціалізованих виданнях, що входять до переліку фахових видань України, в яких публікуються результати досліджень з технічних наук, 1 у електронному фаховому виданні України, 4 у міжнародних зарубіжних періодичних виданнях (1 стаття входить до міжнародних наукометричних баз Scopus та Web of Science), 3 статті у зарубіжних колективних монографіях (3 статті входять до міжнародної наукометричної бази Scopus); 8 праць міжнародних IEEE конференцій (входять до міжнародної наукометричної бази Scopus) та 5 тез доповідей на міжнародних науково-технічних конференціях.

Структура та обсяг роботи. Дисертаційна робота складається з анотації двома мовами, вступу, чотирьох розділів, висновків, списку використаних джерел зі 134 найменувань та 2 додатків; містить 20 таблиць і 27 рисунків та має загальний обсяг 127 сторінок.

ОСНОВНИЙ ЗМІСТ ДИСЕРТАЦІЇ

У **вступі** обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету та завдання дослідження, наукову новизну й практичне

значення отриманих результатів, наведено дані про впровадження результатів роботи, публікації та особистий внесок автора.

Перший розділ дисертації містить аналіз сучасного стану задачі класифікації упорядкованих масивів даних. Під упорядкованим масивом даних кваліфікується множина статистичних даних будь-якої природи, на якій задана функція упорядкування елементів. Найвідомішим представником УМД є часові ряди, які можна тлумачити як реалізації випадкових процесів з дискретним часом. УМД представляють собою дискретну реалізацію деякого випадкового процесу. Самоподібність випадкових процесів полягає в збереженні розподілів ймовірностей при зміні масштабу часу. Стохастичний процес $X(t)$ є самоподібним з параметром H , якщо процес $a^{-H}X(at)$ описується тими ж законами розподілів, що і $X(t)$. Параметр H , $0 < H < 1$, званий показником Херста, являє собою ступінь самоподібності процесу. Поряд з цим показник характеризує міру довгострокової залежності. Якщо $H > 0,5$, процес $X(t)$ є персистентним, тобто трендостійким. У разі $H < 0,5$ процес має антиперсистентні властивості, а при $H = 0,5$ значення процесу є незалежними.

Мультифрактальні об'єкти є статистично неоднорідними самоподібними об'єктами та виявляють більш складну скейлінгову поведінку. У цьому випадку скейлінговою характеристикою є нелінійна функція $h(q)$ – узагальнений показник Херста. Значення $h(q)$ при $q = 2$ збігаються зі значеннями ступеня самоподібності H . Для монофрактальних процесів узагальнений показник Херста $h(q) = H$. Діапазон значень узагальненого показника Херста $\Delta h(q) = h(q_1) - h(q_2)$ визначає ступінь мультифрактальності: чим більше значення $\Delta h(q)$, тим більше виражені мультифрактальні властивості процесу. У разі монофрактальності $\Delta h(q) = 0$. Одним з найбільш затребуваних на практиці методів аналізу самоподібних часових рядів є мультифрактальний детрендований флуктуаційний аналіз (МФДФА). Він дозволяє оцінювати узагальнений показник Херста $h(q)$ для нестационарних часових рядів.

Властивості часових рядів нескладно узагальнити на УМД, якщо замість змінної упорядкування часу взяти іншу фізичну величину. Виходячи з визначення УМД, ІТ повинна класифікувати будь-які дані, які володіють фрактальними властивостями. Наприклад, ІТ може класифікувати дані котирувань обмінних курсів валют, ціни на міжнародних товарних біржах з метою виявлення особливостей в динаміці відповідних котирувань. Також ІТ можна застосовувати для класифікації кардіограм, електроенцефалограм з метою точного виявлення найменших змін в роботі відповідних органів людини, вчасної постановки діагнозу та відповідного лікування. Також ІТ може бути застосована для класифікації сейсмологічних графіків активності земної кори, виявлення DDoS-атак в комп'ютерній мережі та своєчасної їх локалізації, а також для будь-яких інших УМД, котрі володіють фрактальними властивостями.

У **другому розділі** розглянуті та отримали подальший розвиток методи генерації фрактальних УМД, що потрібно для застосування їх у якості навчальної вибірки при роботі з методами машинного навчання. У практичних задачах досить часто виникає проблема недостатнього об'єму даних для навчання моделі машинного навчання, що створює перешкоду для використання ефективних методів розв'язання задачі. У таких ситуаціях зазвичай використовують методи математичного моделювання, котрі дозволяють згенерувати необхідну кількість навчальних прикладів із заданими властивостями – зокрема задача удосконалення методів генерації УМД з заданими фрактальними властивостями є затребуваною та актуальною.

У загальному випадку модельна реалізація, побудована на основі УМД, котрий має фрактальні властивості, може бути представлена наступним чином:

$$Y(p) = f(\text{process}, \text{parameters}, N), \quad \text{parameters} = g(\text{estimates}(X(q), N_x)),$$

де $Y(p)$, $p=1, \dots, N$ – модельна реалізація необхідної довжини N ; *process* – базовий стохастичний фрактальний процес; *parameters* – параметри цього процесу; $X(q)$ – досліджуваний УМД довжини N_x , $q=1, \dots, N_x$; f, g – деякі перетворення, *estimates* – статистичні та фрактальні характеристики, отримані по УМД $X(q)$.

Розглянемо основні моделі УМД, які використовуються в дослідженні в якості навчальної вибірки. Як базові фрактальні процеси було обрано такі: фрактальний броунівський рух (ФБР), експоненціальне перетворення фрактального гаусівського шуму (ЕПФГШ), стохастичний біноміальний мультиплікативний каскад (СБМК) та мультифрактальний броунівський рух (МФБР).

ФБР є самоподібним процесом з єдиним параметром – показником Херста H . ФБР з параметром $H=0,5$ співпадає з класичним броунівським рухом. Прирошення ФБР називаються фрактальним гаусівським шумом (ФГШ). Дисперсія ФГШ відповідає відношенню $D[X(t+\tau) - X(t)] \propto \tau^{2H}$.

В якості випадкового процесу, який породжує модельні реалізації з заданими характеристиками самоподібності та важкістю хвостів, запропоновано використовувати наступне функціональне перетворення ФГШ

$$Y(t) = \text{Exp}[k \cdot X(t)],$$

де $X(t)$ – ряд природжень ФГШ з заданим параметром H , k – параметр, котрий впливає на дисперсію ряду $Y(t)$, тобто на ступінь його фрактальної неоднорідності (мультифрактальності). Таким чином, отримуємо мультифрактальні УМД з заданими показником Херста та мультифрактальними властивостями.

Процедура побудови мультифрактального СБМК складається з двох частин: перша передбачає нескінченну геометричну деталізацію шляхом

ітераційного розбиття інтервалів, а друга забезпечує випадковість вагових коефіцієнтів розбиття, які є значеннями каскаду.

Для побудови СБМК досить часто використовується випадкова величина, що має бета-розподіл $Beta(\alpha, \beta)$. Мультифрактальні властивості реалізації стохастичного каскаду визначаються співвідношенням

$$h(q) = -\frac{1}{q} \log_2 \frac{B(\alpha + q, \beta)}{B(\alpha, \beta)}. \quad (1)$$

Для генерації мультифрактальних реалізацій необхідно перейти від мультифрактальної характеристики $h(q)$ досліджуваного УМД до параметрів бета-розподілу. На практиці найбільш поширені моделі з симетричним бета-розподілом тому, що параметр симетричного бета-розподілу може бути визначеним розв'язанням рівняння (1). У роботі запропонований механізм визначення параметрів α та β несиметричного бета-розподілу, що дозволяє охопити більш великий спектр мультифрактальних властивостей.

Оскільки рівняння (1) не має аналітичного розв'язання відносно невідомих параметрів α та β запропоновано наближене рішення задачі нелінійної умовної оптимізації:

$$\alpha, \beta = \arg \min_{\alpha, \beta} \left(q \cdot h(q) + \log_2 \frac{B(\alpha + q, \beta)}{B(\alpha, \beta)} \right)^2, \quad (2)$$

$$\alpha > 0, \beta > 0, q > -\alpha.$$

Запропонований підхід дозволяє генерувати мультиплікативні біноміальні каскади з довільними мультифрактальними властивостями з діапазоном показника Херста $H \in (0,5; 1,0)$.

Процес МФБР $X(t)$ заснований на моделюванні фрактального броунівського руху в мультифрактальному часі за допомогою операції субординації

$$X(t) = B_H(\theta(t)), \quad t \geq 0,$$

де $\theta(t)$ – біноміальний мультиплікативний каскад, $B_H(t)$ – фрактальний броунівський рух з показником Херста H .

Мультифрактальні властивості реалізацій МФБР $X(t)$ визначаються співвідношенням

$$h_X(q) = H \cdot h_\theta(Hq),$$

де $h_X(q)$ – узагальнений показник Херста для МФБР, $h_\theta(Hq)$ – узагальнений показник Херста для СБМК $\theta(t)$ та H – показник Херста для ФБР $B_H(t)$.

Запропонований підхід до розв'язання оптимізаційної задачі (2) дає змогу моделювати МФБР із заданими властивостями, базуючись на оцінках мультифрактальних характеристик існуючих УМД.

Розглянуті моделі охоплюють широкий спектр фрактальних УМД і можуть бути застосовані для генерації багаточисельних технічних і природних явищ, які мають фрактальні властивості. На рис. 1 показано реалізації розглянутих фрактальних УМД: фрактальний броунівський рух (а), експоненціальне перетворення фрактального гаусівського шуму (б), мультифрактальний броунівський рух (в) та стохастичний біноміальний мультиплікативний каскад (г).

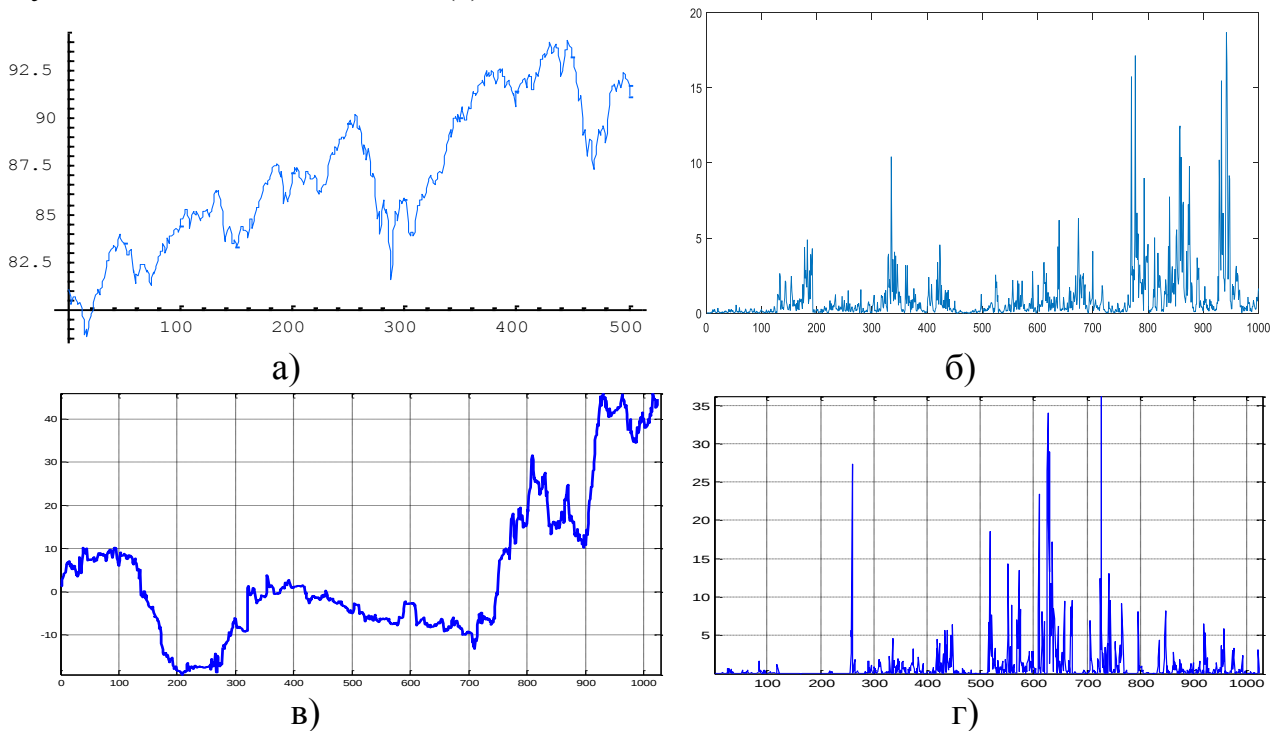


Рисунок 1 – Реалізації: а) ФБР; б) ЕПФГШ; в) МФБР; г) СБМК

У **третьому розділі** представлено опис та результати чисельних експериментів класифікації УМД в залежності від типу УМД, використаних його характеристик для класифікації та методів машинного навчання.

Сформулюємо задачу класифікації УМД таким чином. Є множина УМД, які розділені деяким чином на класи. Визначена скінченна множина УМД, для яких відомо, до яких класів вони належать. Ця множина є навчальною вибіркою. Класова приналежність інших УМД невідома. Необхідно побудувати алгоритм, здатний класифікувати довільний УМД з вихідної множини.

В формальному виді постановка задачі класифікації має такий вигляд. Нехай $x_i \in X, i = \overline{1, n}$ – множина ознак, входів моделі, $y_i \in Y, i = \overline{1, n}$ – множина об'єктів відповідей, виходів моделі. Пара $(x_i, y_i) \in X \times Y$ називається розміченим об'єктом. Комбінація $\{x_i\}, i = \overline{1, n}$ та $\{y_i\}, i = \overline{1, n}$ називається навчальною вибіркою. Задача класифікації полягає у визначенні функції залежності $f: X \rightarrow Y$, котра передбачує по $x \in X$ відповіді $y \in Y$.

Розглянемо підхід до класифікації УМД за їхніми фрактальними властивостями. Змінення фрактальних властивостей УМД безпосередньо пов'язано зі зміною його кореляційної та спектральної структури та ряду

статистичних властивостей, які визначаються специфікою даного УМД. Таким чином, кажучи про деякий клас УМД з близькими фрактальними властивостями, маємо на увазі клас об'єктів із певним комплексом спектрально-кореляційних і статистичних властивостей. Природно за міру поділу на класи за фрактальними властивостями прийняти значення або діапазон значень міри самоподібності об'єкта – показника Херста.

У роботі здійснено ряд чисельних експериментів, в яких проводилася класифікація різних типів УМД, які були розбиті на класи. Алгоритми, розроблені в другому розділі, дали можливість генерувати реалізації УМД, що мають фрактальні властивості з заданим ступенем самоподібності. Кожен клас – це набір згенерованих УМД з показником Херста, що належить заданому діапазону значень. У загальному випадку при генерації реалізацій значення показника Херста змінювалися в інтервалі від H_1 до H_2 з кроком ΔH .

В якості методів машинного навчання використовувалися наступні алгоритми: метод беггінг – з деревами класифікації та регресії; метод випадкового лісу – з деревами класифікації та регресії; нейронні мережі, на основі багатошарового персептрону та рекурентні мережі довго-короткострокової пам'яті (Long Short Term Memory, LSTM) – особливий вид рекурентних нейронних мереж, здатних до навчання довгостроковим залежностям.

Беггінг – технологія класифікації, де всі елементарні класифікатори навчаються та працюють незалежно один від одного. Ідея полягає в тому, що класифікатори не виправляють помилки один одного, а компенсують їх під час голосування. Для отримання результату роботи ансамблю моделей зазвичай використовують наступні способи комбінації: голосування або усереднення. Ефективність методу досягається завдяки тому, що базові алгоритми, навчені на різноманітних підвибірках, є достатньо різноманітними, та їх похибки взаємно компенсуються при голосуванні, а також за рахунок того, що нетипові об'єкти можуть не увійти до деяких навчальних підвибірок.

Випадковий ліс також є методом беггінгу, але, на відміну від його основної версії, має декілька особливостей: використовує лише ансамблі регресійних дерев або дерев класифікації; в алгоритмі семплювання, окрім вибору випадковим чином навчальних об'єктів, також аналогічно здійснюється вибір ознак; для кожної підвибірки дерево рішень будується до повного вичерпання навчальних прикладів та не піддається процедурі відсічення гілок.

Для проведення класифікації в якості нейронних мереж був вибраний багатошаровий персептрон з функцією активації типу ReLu. Ця нейронна мережа є універсальним апроксиматором та здатна виявляти приховані закономірності в даних. При правильному підборі архітектури нейронної мережі навчена модель здатна давати вдалі результати класифікації. Для попередження перенавчання моделі та для збільшення точності класифікації в структуру нейронної мережі включили декілька шарів пакетної нормалізації. Основною відмінністю мережі LSTM від звичайної мережі зі зворотними зв'язками є те, що кожний нейрон прихованого шару LSTM мережі замінений

на комірку пам'яті. Така заміна дозволяє нейронній мережі з LSTM шаром враховувати непостійну періодичність у даних. У мережі LSTM шари використовуються для виявлення закономірностей в даних, але мережа, яка складається лише з LSTM шарів, не здатна класифікувати, тому її необхідно доповнити щільно зв'язаними шарами з функцією активації типу ReLu та шаром пакетної нормалізації.

Для навчання класифікаторів використовувалися декілька різновидів вхідної інформації, а саме: значення УМД, статистичні, фрактальні та рекурентні характеристики УМД. У поданих нижче випадках кожен клас – це набір згенерованих УМД з показником Херста, що належить персистентному діапазону значень. У приведених нижче випадках значення показника Херста змінювалось в інтервалі від 0,5 до 1,0 з кроком 0,05, при цьому мінімальне та максимальне значення були обрані 0,51 та 0,99 відповідно. Таким чином було сформовано 11 класів реалізацій. Величина показника Херста для моделювання реалізації бралася випадково за рівномірним законом розподілу у відповідному інтервалі значень показника Херста. Навчання моделей для кожного класу проводилось на 300 прикладах навчання та перевірялось на 150 тестових. Схема експерименту представлена на рис. 2.

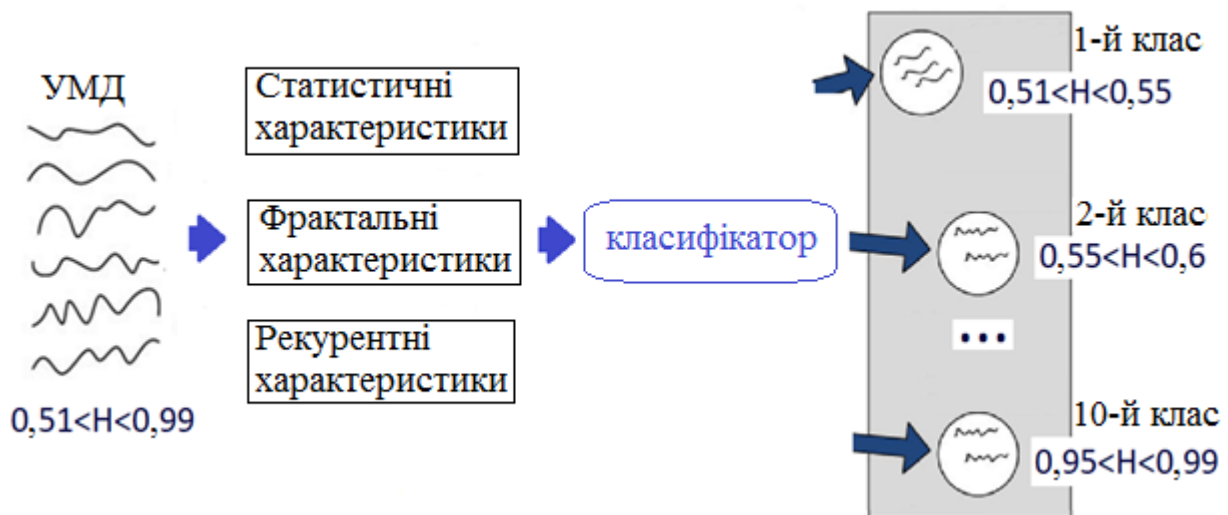


Рисунок 2 – Схематичне зображення класифікації фрактальних реалізацій

Розглянемо результати класифікації, яка була проведена за різними наборами ознак: 1) значення УМД; 2) статистичні та фрактальні характеристики; 3) фрактальні та рекурентні характеристики.

Найбільш простими – у сенсі отримання – ознаками є самі значення УМД. Ці ознаки не потребують ніякого виконання обчислень і перетворень УМД. На вхід класифікатора надходили значення УМД, на виході отримано значення класу, тобто інтервал значень показника Херста, з яким згенерована реалізація.

В табл. 1 представлено середню ймовірність передбачення класу по значеннях УМД для всіх досліджуваних фрактальних моделей в залежності від методу машинного навчання. У цьому випадку довжина УМД складала 1024 значення.

Результати свідчать, що використання значень УМД в якості ознак для класифікації виправдано лише при класифікації фрактальних УМД, які мають сильно виражені мультифрактальні властивості, як, наприклад, реалізації типу біноміального мультиплікативного каскаду. При класифікації реалізацій СБМК доцільно використовувати методи на основі дерев прийняття рішень. У цьому випадку досягається точність більш ніж 0,85. Наприклад, для методу випадкового лісу з деревами класифікації середня ймовірність передбачення класу дорівнює 0,891.

Таблиця 1 – Середня ймовірність передбачення класу, коли ознаки є значеннями УМД

	Беггінг		Випадковий ліс		Нейронні мережи	
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії	Пер-септрон	LSTM
ФБР	0,532	0,484	0,599	0,476	0,302	0,492
ЕПФГШ	0,567	0,498	0,648	0,491	0,292	0,108
СБМК	0,860	0,879	0,891	0,875	0,417	0,415
МФБР	0,447	0,325	0,492	0,324	0,269	0,379

Дослідження показали, що статистичними характеристиками, які відображають зміну фрактальних властивостей УМД, є дисперсія, коефіцієнт варіації, медіана, коефіцієнт асиметрії та ін. Як фрактальні ознаки зручно використовувати значення показника Херста, діапазон значень і деякі значення функції узагальненого показника Херста, отримані методом МФДФА. У цьому випадку на вхід класифікатора надходили оцінки статистичних і фрактальних характеристик, отримані на основі реалізації УМД.

В табл. 2 подано значення середньої ймовірності передбачення класу за статистичними та фрактальними характеристиками для всіх досліджуваних фрактальних моделей в залежності від методу машинного навчання. Довжина реалізацій УМД складала 1024 значення.

Таблиця 2 – Середня ймовірність передбачення класу, коли ознаки є статистичними та фрактальними характеристиками

	Беггінг		Випадковий ліс		Нейронні мережи	
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії	Пер-септрон	LSTM
ФБР	0,154	0,156	0,156	0,155	0,175	0,160
ЕПФГШ	0,969	0,968	0,967	0,969	0,943	0,949
СБМК	0,864	0,866	0,865	0,867	0,755	0,781
МФБР	0,985	0,986	0,982	0,986	0,949	0,957

Результати класифікації показують, що під час використання статистичних та фрактальних характеристик маємо якісну класифікацію для реалізацій всіх моделей, що володіють мультифрактальними властивостями. У даному випадку хорошими класифікаторами зарекомендували себе нейронні мережі. На рис. 3 представлено розподілення ймовірностей визначення номеру класу в залежності від діапазону показника Херста для мультиплікативних каскадів та випадкового лісу з деревами регресії.

Однак, монофрактальні реалізації, куди входить ФБР, практично не розрізняються для малих діапазонів ступеня самоподібності. Щоб домогтися високої точності класифікації для УМД зі слабкими мультифрактальними властивостями, запропоновано використовувати рекурентні діаграми.

Використання рекурентних характеристик в якості ознак в машинному навчанні є новим підходом. Рекурентна діаграма є квадратною матрицею, де елемент з координатами (i, j) характеризує близькість точок i та j УМД в фазовому просторі. Елемент цієї матриці дорівнює 1, якщо відстань у фазовому просторі між точками i та j не перевищує деякої наперед заданої величини, і 0 – в протилежному випадку. Якісний і кількісний аналіз топології рекурентних діаграм дозволяє класифікувати УМД на основі їх статистичних властивостей.

Чисельний аналіз рекурентних діаграм дозволяє обчислювати кількісні міри складності структур рекурентних діаграм, такі як міра рекурентності, міра детермінізму, міра ентропії та ін. Ці характеристики запропоновано застосовувати як ознаки в машинному навчанні.

У цьому випадку на вхід класифікатора надходили оцінки фрактальних і рекурентних характеристик, отримані на основі реалізації УМД: для кожної реалізації методом МФДФА розраховувався вибірковий узагальнений показник Херста і будувалася рекурентна діаграма, на основі якої розраховувався ряд кількісних рекурентних характеристик.

В табл. 3 представлені значення середньої ймовірності передбачення класу за фрактальними та рекурентними характеристиками для всіх досліджуваних фрактальних моделей в залежності від методу машинного навчання. Довжина реалізацій УМД складала 1024 значення.

Таблиця 3 – Середня ймовірність передбачення класу, коли ознаки є фрактальними та рекурентними характеристиками

	Беггінг		Випадковий ліс		Нейронні мережі	
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії	Персептрон	LSTM
ФБР	0,943	0,942	0,855	0,915	0,769	0,964
ЕПФГШ	0,985	0,987	0,985	0,988	0,527	0,489
СБМК	0,921	0,92	0,902	0,921	0,684	0,652
МФБР	0,985	0,987	0,971	0,987	0,792	0,740

Використання рекурентних характеристик разом із фрактальними дозволяє класифікувати УМД зі слабкими мультифрактальними властивостями, або навіть монофрактальні УМД. Також збільшується точність класифікації для УМД з сильно та середньо вираженими мультифрактальними властивостями.

Таким чином, дослідження показали, що оцінювання показника Херста доцільно виконувати методами машинного навчання за значеннями УМД, за значеннями рекурентних та фрактальних характеристик в залежності від фрактальних властивостей УМД. Запропонований підхід дозволяє зменшити довірчий інтервал оцінки показника Херста в декілька разів. У табл. 4 представлені усереднені оцінки ширини 95% довірчого інтервалу для різних типів УМД при оцінюванні методом МФДФА та методами машинного навчання.

Таблиця 4 – Ширина довірчого інтервалу, визначеного МФДФА та методами машинного навчання

Довжина УМД	Довірчий інтервал оцінки H	
	Машинне навчання	МФДФА
256	0,13	0,35
512	0,11	0,24
1024	0,09	0,19

У **четвертому розділі** розроблені методологія та ІТ класифікації фрактальних упорядкованих масивів даних. Розроблена UML-модель запропонованої ІТ класифікації фрактальних упорядкованих масивів даних наведена на рис. 4, 5.

Розглянемо короткий опис запропонованої ІТ.

1. Збереження даних до бази даних (БД). Цей модуль ІТ забезпечує збереження вхідного потоку інформації до БД для можливості повторного використання.

2. Обробка інформації. Представляє собою комплекс математичного апарату для перетворення та генерації даних:

2.1. Попередня обробка даних. Під цим пунктом розуміється досить великий спектр алгоритмів, що направлені на виявлення апріорно відомої інформації про природу УМД, його фрактальних і статистичних властивостей.

2.2. Оцінювання самоподібних та мультифрактальних властивостей УМД для різних класів. Даний крок є найбільш важливим для вибору методу класифікації. Ключовим моментом є визначення факту, що різні класи УМД мають досить різні фрактальні властивості. Якщо фрактальні властивості для різних класів практично однакові, то даний підхід не має сенсу використовувати.

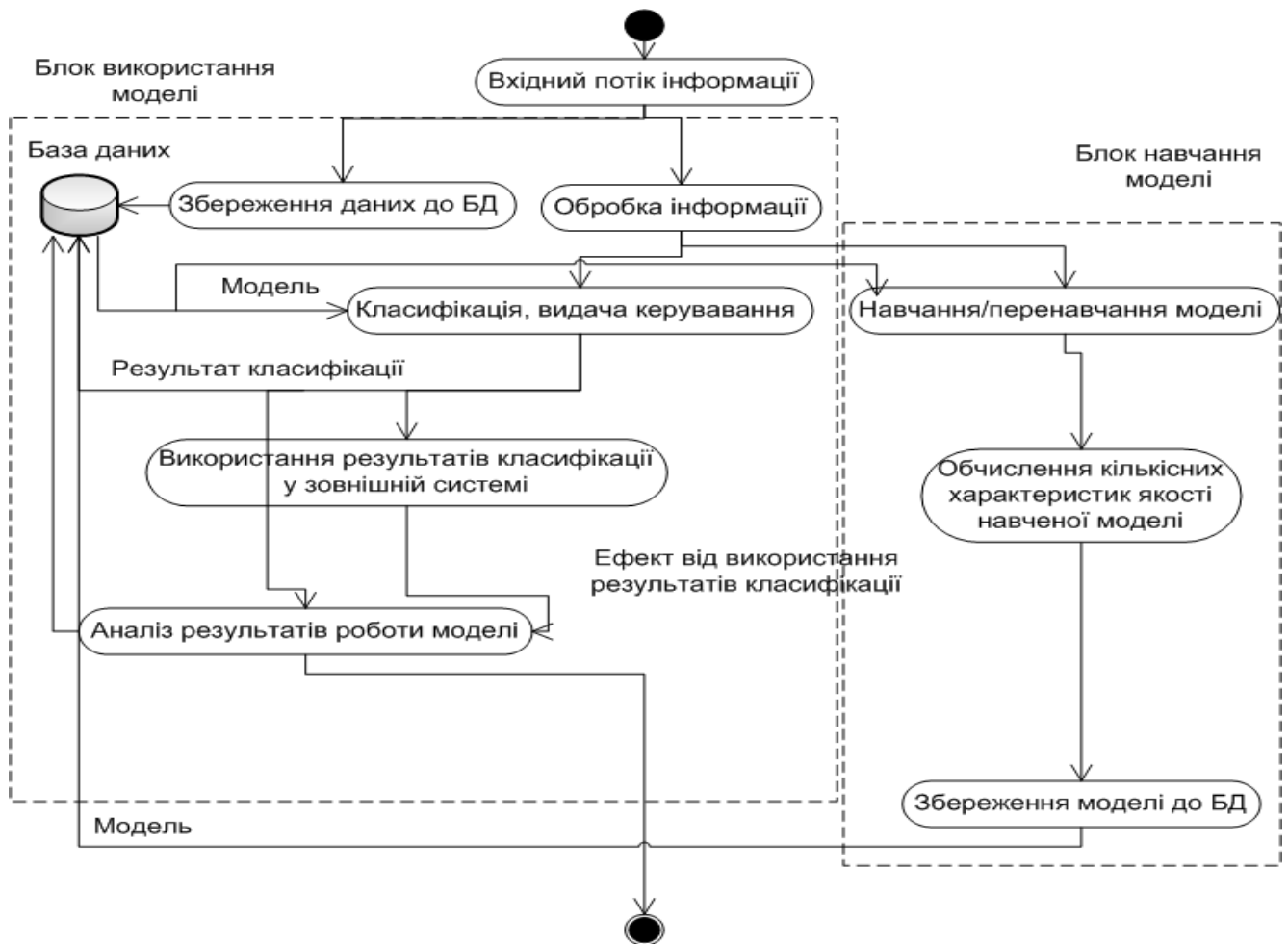


Рисунок 4 – Схема ІТ класифікації упорядкованих масивів даних

2.3. Аналіз самоподібних властивостей УМД дозволяє виділити три умовних діапазони для показника Херста H : антиперсистентність при , персистентність при значеннях і фактична відсутність довгострокової залежності . У даному випадку, враховуючи похибки оцінювання показника Херста, приймаємо, що діапазон значень $[0,45; 0,55]$ відповідає дуже малій автокореляційній залежності.

2.4. Аналіз мультифрактальних властивостей УМД, тобто оцінювання узагальненого показника Херста, дозволяє виділити три умовних діапазони мультифрактальності для функції $h(q)$: слабкі мультифрактальні властивості $\Delta h(q) < 0,4$, середні мультифрактальні властивості $0,4 \leq \Delta h(q) \leq 1,0$ і сильну мультифрактальність $\Delta h(q) > 1,0$. Керуючись дослідженнями, зроблено висновок, що діапазон мультифрактальних і самоподібних властивостей УМД відіграє важливе значення при виборі класифікатора і, відповідно, точності класифікації. Дослідження показали, що найбільш добре піддаються класифікації УМД з сильними мультифрактальними властивостями. У цьому випадку і персистентні і антиперсистентні УМД найбільш точно класифікуються за допомогою метода випадкового лісу на основі дерев регресії (ВЛ[ДР]). В якості ознак достатньо використовувати статистичні (С) та фрактальні (Ф) характеристики УМД. Якщо УМД мають трендостійкість

(показник Херста $H > 0,5$), то в якості ознак використовуються самі значення УМД. Це займає набагато більше часу, але не потребує ніякого оцінювання фрактальних характеристик.

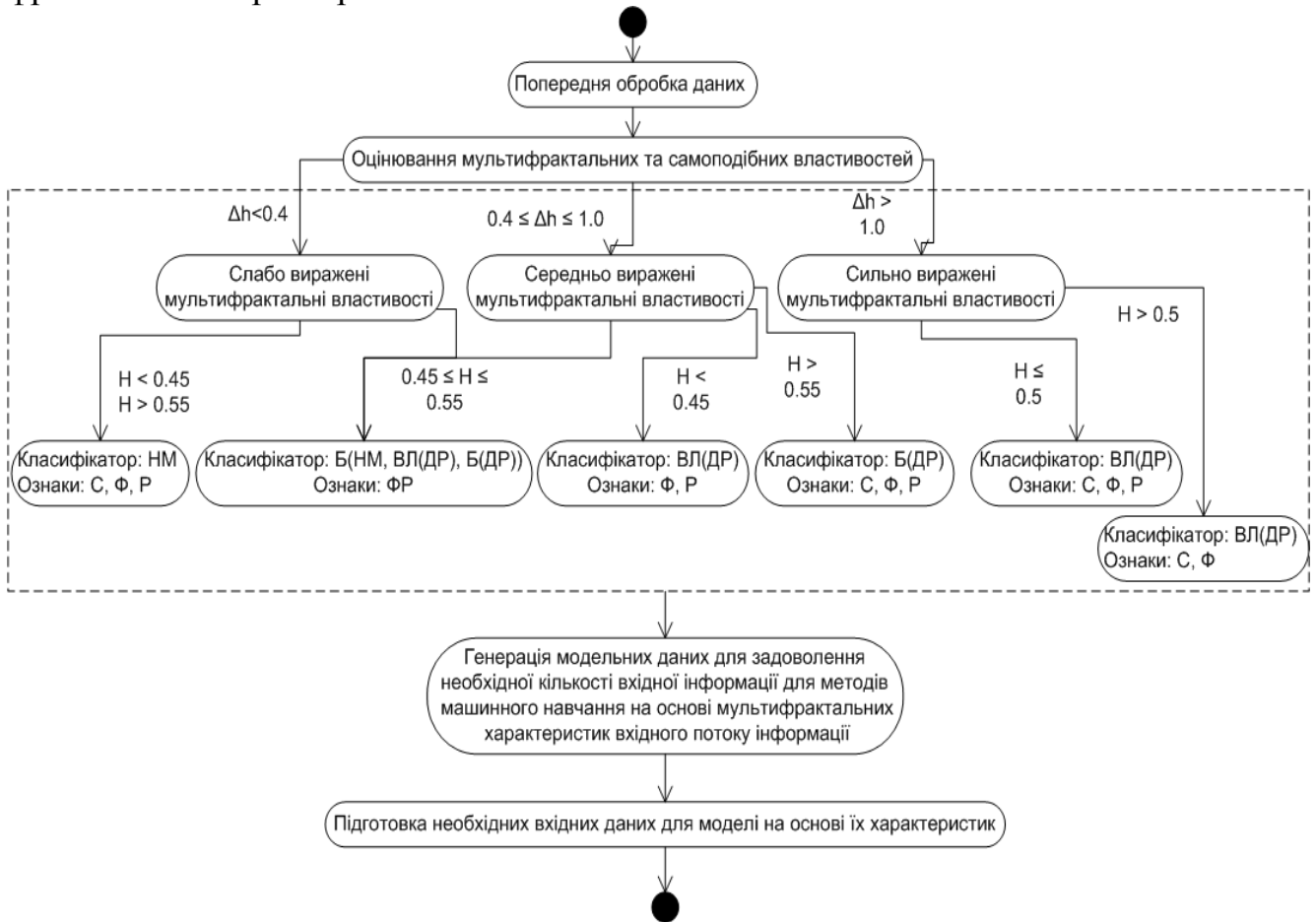


Рисунок 5 – Декомпозиція блоку «Обробка інформації»

2.5. У разі, коли УМД явно має мультифрактальні властивості, але діапазон узагальненого показника Херста має деяке невелике значення $0,4 \leq \Delta h(q) \leq 1,0$, в якості ознак також можна застосовувати статистичні та фрактальні характеристики. Причому в персистентному випадку краще працює метод беггінг з деревами регресії (Б[ДР]), а в антиперсистентному – метод випадкового лісу із деревами регресії (ВЛ[ДР]).

2.6. Якщо мультифрактальні властивості УМД слабо виражені, тобто УМД вважаються умовно монофрактальними, то в якості додаткових ознак треба використовувати рекурентні (Р) характеристики. Дослідження показали, що у випадку монофрактальності кращі результати класифікації дає застосування нейронних мереж (НМ).

2.7. Найбільш складним випадком класифікації є варіант з класифікацією УМД, які мають слабку та середню мультифрактальність, та у яких практично відсутня автокореляційна залежність ($0,45 \leq H \leq 0,55$). У цьому випадку пропонується застосувати ансамбль (беггінг) з декількох операторів, а саме Б{НМ, ВЛ[ДР], Б[ДР]}.

2.8. Блок генерації модельних даних, необхідний для використання ІТ в умовах недостатньої кількості інформації.

3. Навчання/перенавчання моделі. Цей блок ІТ відповідає за навчання та періодичне перенавчання моделі для класифікації УМД.

4. Обчислення кількісних характеристик якості навченої моделі. Відповідає за обчислення статистичних оцінок якості отриманої моделі, даний модуль необхідний для розуміння ступеня точності моделі.

5. Збереження моделі та її характеристик до БД. Відповідає за збереження нової версії моделі до бази даних.

6. Класифікація, видача управління. Цей модуль відповідає за використання актуальної версії моделі для класифікації вхідного потоку даних та генерації керуючого сигналу на основі результатів класифікації. Якщо модель ще не навчена, то цей та наступні модулі свої функції не виконують до моменту завершення збору необхідного об'єму інформації та навчання моделі.

7. Використання результатів класифікації. Використання сигналу управління зовнішньою системою та отримання відповіді на управління.

8. Аналіз результатів роботи моделі. Цей модуль відповідає за порівняння запропонованого управління та очікуваного результату його використання з фактичним відгуком зовнішньої системи.

9. Збір та генерація статистичних оцінок результатів використання ІТ. Цей блок відповідає за усереднення результатів роботи ІТ та генерації звітів.

Також у розділі наведено приклад практичного застосування ІТ класифікації. Розглянемо застосування запропонованої технології на прикладі детектування DDoS-атак – хакерських атак на обчислювальну систему з метою створення таких умов, при яких користувачі системи не можуть отримати доступ до системних ресурсів. У даний час DDoS-атаки дозволяють довести до відмови практично будь-яку систему, не залишаючи юридично значимих доказів.

Одним із рішень задачі своєчасного виявлення атаки є розробка класифікатора, який визначає ймовірність того, що аналізований трафік є атаківаним. Останні дослідження показують, що однією з характерних ознак атаки є зміна показника Херста у трафіку, що містить атакуючі файли.

Для проведення експериментів були взяті дані трафіків з двох наборів реальних даних. З першого набору були взяті реалізації DDoS-атак, а з другого – дані трафіку реальної мережі Internet Service Provider на каналному рівні. Реалізація трафіку під дією DDoS-атаки є сумою трафіку та реалізації одного з видів атаки. У ході роботи проведені експерименти по обчисленню зміни узагальненого показника Херста під дією DDoS-атак.

У табл. 5 представлені фрактальні властивості для різних нормальних і атаківаних трафіків. Очевидно, що під дією атаки мультифрактальні властивості посилюються, а показник Херста збільшується. Це пов'язано з тим, що типові часові реалізації DDoS-атак мають великий показник Херста та сильні мультифрактальні властивості.

Таблиця 5 – Фрактальні властивості нормальних і атакованих трафіків

		$h(0,1)$	$h(5)$	H	$\Delta h(q)$
1.	Трафік	2,2	0,4	0,57	1,8
	Атакований трафік	2,8	0,41	0,61	2,39
2.	Трафік	1,78	0,24	0,51	1,54
	Атакований трафік	2,05	0,26	0,6	1,79
3.	Трафік	1,8	0,2	0,53	1,27
	Атакований трафік	2,5	0,3	0,64	1,86

Для оцінки якості роботи ІТ була розгорнута віртуальна мережа, в якій трафік передавався з п'яти джерел на сервер ідентифікації атаки. На окремому сервері була розгорнута запропонована ІТ та проводилося навчання моделей класифікації з подальшим їх використанням для визначення атак в мережі. Увесь трафік в мережі відстежувався і оброблявся на сервері як в режимі реального часу, так і шляхом автономної обробки раніше захопленого мережного трафіку. Це дозволяє аналізувати трафік з різних джерел різної інтенсивності і розставляти пріоритети для його обробки.

У табл. 6 представлено показники якості роботи запропонованої ІТ та стандартних методів виявлення вторгнень. Для аналізу якості роботи запропонованої ІТ ідентифікації атак використовувались такі показники ефективності: відсоток виявлених атак, відсоток атак, які не були виявлені, відсоток втрачених даних від загальної кількості отриманих даних, відсоток помилкових спрацьовувань.

Таблиця 6 – Якість роботи запропонованої інформаційної технології

	Виявлення атак з використанням ІТ	Стандартні методи виявлення атак без використання ІТ
% виявлених атак	97,2	94,3
% невиявлених атак	0,3	0,5
% втрачених даних	3,0	3,0
% помилкових спрацьовувань	1,5	1,6

Під час порівняння якості роботи запропонованої ІТ для виявлення вторгнень з існуючими методами визначено, що запропонований метод краще ідентифікує атаки, при цьому кількість втрачених даних і помилкових спрацьовувань є незначно меншими та час ідентифікації атак відносно однаковий при роботі існуючих методів.

У **додатках** наведено акти впровадження розроблених методів.

ВИСНОВКИ

У дисертаційній роботі вирішена важлива науково-практична задача розробки ІТ класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання з метою підвищення точності класифікації для аналізу та діагностування впорядкованих даних різної природи. Результати проведених досліджень полягають у наступному.

1. Проведено аналіз існуючих методів класифікації упорядкованих масивів даних методами машинного навчання та виявлено шляхи збільшення точності класифікації.

2. Подальший розвиток отримав метод генерації мультифрактальних реалізацій на основі стохастичного мультиплікативного біноміального каскаду, що дозволяє генерувати упорядковані масиви даних із заданими мультифрактальними характеристиками для стохастичного біноміального каскаду та мультифрактального броунівського руху.

3. Здійснено чисельні експерименти класифікації різних типів упорядкованих даних, при цьому множина УМД розбивалась на класи за їхніми фрактальними властивостями. Як класифікатори були застосовані методи дерев рішень та нейронних мереж. Дослідження показали, що діапазон мультифрактальних і самоподібних властивостей УМД відіграє важливе значення для вибору класифікатора і, відповідно, точності класифікації. Запропоновано використовувати рекурентні характеристики УМД як ознаки для класифікації методами машинного навчання – це дозволило збільшити точність класифікації в загальних випадках та дало змогу класифікувати УМД з монофрактальними властивостями. Запропоновано метод оцінювання показника Херста за значеннями УМД, який засновано на застосуванні класифікаційних методів машинного навчання, що дозволяє зменшити довірчий інтервал оцінки показника Херста в декілька разів.

4. Розроблено інформаційну технологію класифікації фрактальних упорядкованих масивів даних, яка заснована на поєднанні методів дерев рішень та нейронних мереж, враховує самоподібні та мультифрактальні властивості упорядкованих даних, що дозволяє збільшити точність класифікації.

5. Результати дисертаційного дослідження впроваджені на підприємстві ПрАт «Фарлеп-Інвест» (м. Харків) для зменшення негативного впливу кібератак на інформаційні системи підприємства, на підприємстві КБ «ПриватБанк» (м. Харків) для виявлення вторгнень в інформаційні системи підприємства та у навчальному процесі в Харківському національному університеті радіоелектроніки при викладанні дисципліни «Artificial Intelligence (Штучний інтелект)».

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Булах В.А. Моделирование мультипликативных стохастических каскадов на основе случайной величины с бета распределением. *International journal information content & processing*. Bulgaria, 2017. Vol. 4, No 2. P. 159–170.

2. Кириченко Л.О., Булах В.А., Радивилова Т.А. Классификация мультифрактальных стохастических временных рядов с использованием мета-алгоритмов на основе деревьев решений. *Системні технології*. Дніпро, 2018. № 3(116). С. 22–27.

Здобувачем розроблений метод класифікації фрактальних часових рядів.

3. Кириченко Л., Булах В., Радивилова Т., Черняк В. Анализ взаимозависимости временных рядов биткоина и активности сообществ в социальных сетях. *International journal information technologies & knowledge*. Bulgaria, 2018. Vol. 12, No 1. P. 43–55.

Здобувачем проведено мультифрактальних аналіз часових рядів біткоїна та активності в соціальних мережах.

4. Kirichenko L., Radivilova T., Bulakh V. Generalized approach to Hurst exponent estimating by time series. *Informatyka automatyka pomiaru w gospodarce i ochronie srodowiska*. Poland, 2018. Vol. 8, No 1. P. 28–31.

Здобувачем проведено дослідження узагальненого підходу до оцінки показника Херста за часовими рядами.

5. Kirichenko L., Bulakh V., Radivilova T. Machine learning in classification time series with fractal properties. *Data*. 2019. Vol. 4, No 5. P. 1–13.

Здобувачем запропоновано методи класифікації часових рядів на основі фрактальних властивостей.

6. Радивилова Т., Кириченко Л., Булах В. Обнаружение DDoS-атак методами машинного обучения на основе фрактальных свойств. *Security in cyberspace, the social internet space in context values and hazards*. Slupsk-Charkow, Poland, 2019. P. 299–315.

Здобувачем досліджено мультифрактальні властивості та запропоновано методи класифікації часових рядів.

7. Kirichenko L., Radivilova T., Bulakh V. Binary classification of fractal time series by machine learning methods. *Lecture notes in computational intelligence and decision making. Intellectual systems of decision making and problems of computational intelligence: proceedings of the XV international scientific conference*. May 21–25, 2019. Springer, Cham, 2020. P. 701–711.

Здобувачем проведено порівняльний аналіз методів класифікації нейронними мережами та деревами прийняття рішень.

8. Кіріченко Л.О., Булах В.А., Тавалбех М.Ф., Зінченко П.П. Інформаційна технологія класифікації фрактальних часових рядів. *Системні технології*. 2020. № 3 (128). С. 115–126.

Здобувачем розроблена та запропонована інформаційна технологія класифікації фрактальних часових рядів.

9. Булах В.А. Інформаційна технологія класифікації упорядкованих масивів даних з фрактальними властивостями. *Радіоелектроніка та інформатика*. Харків, 2020. № 2 (89). С. 66–72.

10. Кіріченко Л.О., Булах В.А., Радівілова Т.А., Тавалбех М.Ф., Зінченко П.П. Балансування самоподібного трафіку в мережних системах виявлення вторгнень. *Кібербезпека: освіта, наука, техніка*. Київ, 2020. №3 (7). С. 17–32. URL: <https://csecurity.kubg.edu.ua/index.php/journal/article/view/130>.

Здобувачем запропоновано використовувати методи машинного навчання для виявлення факту вторгнення в комп'ютерну мережу.

11. Radivilova T, Bulakh V., Ageiev D., Kirichenko L. The methods to improve quality of service by accounting secure parameters. *Advances in computer science for engineering and education*, Springer. Cham, 2020. Vol. 938, P. 346–355.

Здобувачем проведено дослідження методів виявлення вторгнень в комп'ютерні мережі.

12. Kirichenko L., Bulakh V., Radivilova T. Fractal time series analysis of social network activities. *Problems of infocommunications. Science and technology: proceedings of the 4th international scientific-practical conference*. October 10–13, 2017. Kharkiv, Ukraine, 2017. P. 456–459.

Здобувачем досліджено взаємозв'язок рядів біткоїна та активності соціальних мереж методами фрактального аналізу.

13. Кіріченко Л.О., Радівілова Т.А., Булах В.А. Мультифрактальний аналіз часових рядів в соціальних мережах. *Практичне застосування нелінійних динамічних систем в інфокомунікаціях: матеріали 6 міжнародної науково-практичної конференції (1 міжнародний симпозіум)*. Чернівці, 2017. С. 19–20.

Здобувачем проведено мультифрактальний аналіз та досліджено взаємозв'язок рядів біткоїна та активності соціальних мереж.

14. Radivilova T., Kirichenko L., Bulakh V., Ageyev D., Tawalbeh M. Decrypting SSL/TLS traffic for hidden threats detection. *Dependable systems, services and technologies: proceedings of the IEEE 9th international conference*. May 24–27, 2018. Kyiv, Ukraine, 2018. P. 143–146.

Здобувачем проведено дослідження методів виявлення вторгнень в комп'ютерні мережі.

15. Bulakh V. Kirichenko L., Radivilova T. Classification of multifractal time series by decision tree methods. *ICT in education, research and industrial applications. Integration, harmonization and knowledge transfer: proceedings of the 14th international conference*. May 14–17, 2018. Kyiv, Ukraine, 2018. Vol. I (2105). P. 457–460.

Здобувачем проведено порівняльний аналіз методів класифікації нейронними мережами та деревами прийняття рішень.

16. Bulakh V., Kirichenko L., Radivilova T. Time series classification based on fractal properties. *Data stream mining & processing: proceedings of the IEEE second international conference*. August 21–25, 2018. Lviv, Ukraine, 2018. P. 198–201.

Здобувачем проведено порівняльний аналіз класифікації часових рядів заснованої на фрактальних характеристиках.

17. Kirichenko L., Bulakh V., Radivilova T. Classification of fractal time series using recurrence plots. *Problems of infocommunications. Science and technology: international scientific-practical conference*. October 9–12, 2018. Kharkiv, Ukraine, 2018. P. 719–724.

Здобувачем проведено порівняльний аналіз класифікації часових рядів заснованої на рекурентних характеристиках.

18. Kirichenko L., Bulakh V., Radivilova T., Kobitska Y. Fractal analysis of DDoS attack realizations. *Фізико-технічні проблеми передавання, обробки та зберігання інформації в інфокомунікаційних системах: Матеріали 7-ої міжнародної науково-практичної конференції*. 8-10 листопада 2018 р. Чернівці, 2018. С. 13.

Здобувачем проведено фрактальний аналіз DDoS-атак.

19. Radivilova T., Bulakh V., Kirichenko L. The method to control buffering and bandwidth network by monitoring of fractal traffic and security parameters. *Фізико-технічні проблеми передавання, обробки та зберігання інформації в інфокомунікаційних системах: матеріали 7-ої міжнародної науково-практичної конференції*. Чернівці, 2018. С.97.

Здобувачем досліджено мультифрактальні властивості та запропоновано методи класифікації часових рядів.

20. Кириченко Л.О., Булах В.А., Радивилова Т.А. Сравнительный анализ классификации мультифрактальных временных рядов. *Інформаційні технології в металургії та машинобудуванні: матеріали міжнародної науково-технічної конференції*. 27 – 29 березня 2018 р. Дніпро, 2018. С.126.

Здобувачем запропоновано використання фрактальних, рекурентних характеристик часових рядів як ознак при машинному навчанню.

21. Radivilova T., Kirichenko L., Ageyev D., Bulakh V. Classification methods of machine learning to detect DDoS attacks. *Intelligent data acquisition and advanced computing systems: technology and applications: proceedings of the 10th IEEE international conference*. September 18–21, 2019. Metz, France, 2019. P. 207–210.

Здобувачем проведено фрактальний аналіз DDoS-атак, запропоновано метод виявлення DDoS-атак заснований на машинному навчанні.

22. Radivilova T., Kirichenko L., Bulakh V. Comparative analysis of machine learning classification of time series with fractal properties. *Advanced optoelectronics and lasers: proceedings of the IEEE 8th international conference*. September 06–08, 2019. Sozopol, Bulgaria, 2019. P. 557–560.

Здобувачем запропоновано використання фрактальних, рекурентних характеристик часових рядів як ознак при машинному навчанню.

23. Кириченко Л.О., Булах В.А. Классификация временных рядов на основе фрактальных свойств. *Міжнародний науковий симпозіум Інтелектуальні рішення: матеріали V-ої міжнародної науково-практичної конференції*. 15–20 квітня 2019р. Ужгород, 2019. С. 220.

Здобувачем проведено порівняльний аналіз класифікації фрактальних процесів.

24. Radivilova T., Kirichenko L., Bulakh V. Machine learning classification of multifractional brownian motion realizations. *Computer modeling and intelligent systems: proceedings of the third international workshop*. April 27–May 1, 2020. Zaporizhzhia, Ukraine, 2020. P. 980–989.

Здобувачем промодельовано мультифрактальні часові ряди.

АНОТАЦІЇ

Булах В.А. Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання. На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Національний технічний університет «Харківський політехнічний інститут», Харків, 2021 р.

Дисертацію присвячено розробці інформаційної технології класифікації упорядкованих масивів даних (УМД), які мають фрактальні властивості, з використанням методів машинного навчання.

В роботі програмно реалізовано методи генерації УМД з мультифрактальними властивостями різних типів. Здійснено чисельні експерименти, при виконанні яких проводилася класифікація різних типів упорядкованих даних, множина УМД розбивалась на класи за їхніми фрактальними властивостями. Як класифікатори застосовано ансамблеві методи дерев рішень та нейронні мережі, у якості ознак при класифікації використовувалися статистичні, фрактальні та рекурентні характеристики УМД.

Дослідження показали, що діапазон мультифрактальних і самоподібних властивостей масивів даних відіграє важливе значення для вибору класифікатора і набору ознак, та, відповідно, точності класифікації. Якщо впорядковані дані характеризуються сильно вираженими мультифрактальними характеристиками, в більшості випадків достатньо використовувати значення УМД як ознаки при класифікації з використанням ансамблевих методів дерев рішень; також вони ефективно класифікуються за фрактальними характеристиками. Якщо УМД має монофрактальні властивості, то такі дані доцільно класифікувати з використанням рекурентних та фрактальних характеристик за допомогою нейронних мереж. Найбільш складним випадком є класифікація УМД, які мають слабо виражену мультифрактальність та слабку автокореляційну залежність. У цьому випадку пропонується застосувати ансамбль з використанням як окремих класифікаторів нейронних мереж та випадкового лісу, де в якості ознак використовуються фрактальні та рекурентні характеристики УМД.

Розроблена технологія дозволяє використовувати її для обчислення показника Херста за часовими рядами та дозволяє зменшити довірчий інтервал

оцінки показника Херста в декілька разів.

Запропонована інформаційна технологія аналізує вхідний потік інформації та обирає набір характеристик класифікатора для максимізації точності класифікації УМД. Таким чином, інформаційна технологія дозволяє класифікувати дані з різними фрактальними властивостями, що дозволяє використовувати її для класифікації упорядкованих масивів даних різноманітної природи, наприклад, для виявлення DDoS-атак в інфокомунікаційних даних, уточнення діагнозу за записами електроенцефалограми та електрокардіографії, класифікації сейсмічних подій за сейсмограмами тощо.

Ключові слова: класифікація, інформаційна технологія, машинне навчання, рекурентна діаграма, мультифрактальний процес, випадковий ліс, нейронна мережа, мультифрактальний аналіз.

Булах В.А. Информационная технология классификации упорядоченных массивов данных с фрактальными свойствами методами машинного обучения. На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Национальный технический университет «Харьковский политехнический институт», Харьков, 2021 г.

Диссертация посвящена разработке информационной технологии классификации упорядоченных массивов данных (УМД), которые имеют фрактальные свойства, с использованием методов машинного обучения.

В работе программно реализованы методы генерации УМД с мультифрактальными свойствами различных типов. Выполнен ряд экспериментов, в которых проводилась классификация различных типов упорядоченных данных, разбитых на классы по их фрактальным свойствам. Как классификаторы были применены ансамблевые методы деревьев решений и нейронных сетей, в качестве признаков при классификации использовались статистические, фрактальные и рекуррентные характеристики УМД.

Исследования показали, что диапазон мультифрактальных и самоподобных свойств массивов данных играет важное значение для выбора классификатора и набора признаков, и, соответственно, точности классификации. Предложенная информационная технология анализирует входной поток информации и выбирает набор характеристик классификатора для максимизации точности классификации. Разработанная технология позволяет классифицировать данные с различными фрактальными свойствами, например, для выявления DDoS-атак в инфокоммуникационных данных, уточнения диагноза по записям электроэнцефалограммы и электрокардиографии, классификации сейсмических событий по сейсмограммам и тому подобное.

Ключевые слова: классификация, информационная технология, машинное обучение, рекуррентная диаграмма, мультифрактальный процесс, случайный лес, нейронная сеть, мультифрактальный анализ.

Bulakh V. Information technology of classification of ordered data arrays with fractal properties using machine learning methods. As a manuscript.

Thesis for the degree of candidate of technical sciences in specialty 05.13.06 – information technologies. – National Technical University "Kharkiv Polytechnic Institute", Kharkiv, 2021.

The thesis is devoted to the development of information technology for the classification of ordered data arrays (ODA), which have fractal properties, using machine learning methods.

The methods for generating ODA with multifractal properties of various types are implemented in software. The series of experiments in which different types of ordered data split into classes according to their fractal properties were carried out. Ensemble methods of decision trees and neural networks were used as classifiers. The statistical, fractal and recurrence characteristics of ODA were used as features in the classification.

Studies have shown that the range of multifractal and self-similar properties of data set plays an important role in the choice of a classifier and set of features, and, accordingly, classification accuracy. The proposed information technology analyzes the input information flow and selects classifier and features r to maximize the classification accuracy. The developed technology makes it possible to classify data with various fractal properties, for example, to detect DDoS attacks in infocommunication data, to clarify the diagnosis based on electroencephalogram and electrocardiography records, to classify seismic events for seismograms, etc.

Keywords: classification, information technology, machine learning, recurrent diagrams, multifractal processes, random forest, neural networks, multifractal analysis.

