ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

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

Кваліфікаційна наукова
праця на правах рукопису

БУЛАХ ВІТАЛІЙ АНАТОЛІЙОВИЧ

УДК 004.9; 519.2

ДИСЕРТАЦІЯ

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

Спеціальність 05.13.06 – інформаційні технології
Галузь знань 12 – інформаційні технології

Подається на здобуття наукового ступеня
кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ В.А. Булах

Науковий керівник
Кіріченко Людмила Олегівна
доктор технічних наук, професор

Ідентичність копії змісту з першим примірником дисертації засвідчую. Вчений секретар спеціалізованої вченої ради К 64.050.07 Юрій ДОРОФЕЄВ 10.02.2021

Харків – 2021

АНОТАЦІЯ

Булах В.А. Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – «Інформаційні технології». – Міністерство освіти і науки України національний технічний университет «Харківський політехнічний інститут», 2021.

Дисертаційна робота присвячена вирішенню задачі класифікації фрактальних упорядкованих масивів даних (УМД) в різноманітних сферах людської діяльності, наприклад в медицині при діагностиці захворювань по енцефалограмам, кардіограмам; при виявленні кризових ситуацій по біржовим індексам; виявлення вторгнень в комп'ютерну мережу та інше.

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

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.

Проведений у роботі аналіз існуючих методів класифікації упорядкованих масивів даних показав: незважаючи на те, що застосування машинного навчання в задачах аналізу фрактальних даних постійно зростає, на сьогоднішній, день не розроблено інформаційної технології, яка призначена для задач класифікації фрактальних масивів даних, зокрема часових рядів.

У роботі виконано огляд існуючих методів генерування фрактальних та мультифрактальних упорядкованих масивів даних. Доведена необхідність удосконалення таких методів, та запропоновано підхід до генерації стохастичного мультиплікативного біноміального каскаду на основі вирішення задачі нелінійної оптимізації. Розв’язання цієї задачі дозволяє використовувати несиметричний бета-розподіл для генерації стохастичного мультиплікативного біноміального каскаду, що дозволило охопити широкий спектр мультифрактальних властивостей. Використання удосконаленого методу генерації стохастичного мультиплікативного біноміального каскаду дає змогу генерувати більш складні у фрактальному сенсі реалізації мультифрактального броунівського руху.

Проведено чисельні експерименти класифікації різних типів упорядкованих масивів даних, які були розбиті на класи за їхніми фрактальними властивостями. Були застосовані наступні методи класифікації: дерева прийняття рішень, bagging, випадковий ліс, багатошаровий персептрон, мережі довгої короткострокової пам’яті. Дослідження показали, що діапазон мультифрактальних і самоподібних властивостей відіграє важливу роль для вибору класифікатора, і відповідно на точність класифікації. Було доведено що використання статистичних та фрактальних характеристик мультифрактальних упорядкованих масивів даних дозволяє покращиті точність класифікації УМД порівняно з використанням значень УМД як ознак при класифікації.

Розглянуто рекурентні діаграми та запропоновано використання їх характеристик в якості ознак при навчанні моделей класифікації. Дослідження показали що характеристики рекурентних діаграм дають суттєвий внесок в точність класифікації, та дозволяють класифікувати УМД з монофрактальними властивостями.

Розроблена інформаційна технологія класифікації УМД, котра дозволяє підібрати оптимальний метод машинного навчання для класифікації досліджуваного упорядкованого масиву даних.

Запропоновано використання розробленої інформаційної технології для визначення показника Херста УМД. Експериментально доведено ефективність використання методів машинного навчання для визначення показника Херста та його довірчого інтервалу, та дозволило скоротити довірчий інтервал показника Херста для фрактальних та мультифрактальних УМД більше ніж у 2 рази.

Приведено результати застосування запропонованої інформаційної технології для виявлення DDoS-атак в комп'ютерній мережі. ІТ збільшує кількість виявлених атак, що дозволяє використовувати запропоновану ІТ як додатковий фільтр при виявленні DDoS-атак.

Запропоновані у роботі моделі, методи та інформаційна технологія класифікації фрактальних даних були впроваджені у діяльність підприємств різних галузей економіки:

– для підприємства ПрАт «Фарлеп-Інвест» (м. Харків) впроваджено розроблену інформаційну технологію для зменшення негативного впливу кібер-атак на інформаційні системи підприємства;

– для підприємства КБ «ПриватБанк» (м. Харків) впроваджено розроблену інформаційну технологію для виявлення вторгнень в інформаційні системи підприємства.

Результати дисертаційного дослідження впроваджені у навчальному процесі Харківському національному університеті радіоелектроніки на кафедрі інфокомунікаційної інженерії при викладанні навчальної дисципліни «Artificial Intelligence (Штучний інтелект)».

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

Список публікацій здобувача

Наукові праці, в яких опубліковано основні наукові результати 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.

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.

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

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.

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.

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.

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.

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.

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.

Матеріали та праці наукових конференцій:

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

22. Radivilova T., Kirichenko L., Bulakh V. Comparative analysis of machine learning classification of time series with fractal properties. *Advanced optoelectronics*

*and lasers: p*roceedings of the *IEEE 8th international conference.* September 06–08, 2019. Sozopol, Bulgaria, 2019. P. 557–560.

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.

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.

ABSTRACT

Bulakh V.A. Information technology of classification of ordered data sets with fractal properties by machine learning methods. – Qualifying scientific work on the rights of the 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 problem of classification of fractal ordered data sets (ODS) in various spheres of human activity, for example in medicine at diagnosis of diseases on encephalograms, cardiograms; when identifying crisis situations on stock exchange indices; to detect computer network intrusions and more.

The purpose of the thesis is to increase the classification accuracy of ordered data sets with fractal properties by creating information technology of classification by machine learning methods.

The object of research is the process of classifying objects by machine learning methods.

The subject of research - methods and models of classification for fractal ordered data sets.

Research methods. When solving the tasks, the following methods were used: mathematical apparatus of machine learning; methods of mathematical statistics for statistical analysis of ordered data sets and analysis of classification accuracy using machine learning; methods of chaos theory and fractal analysis for calculating fractal and recurrence characteristics..

The scientific novelty of the obtained results is that for the first time the following was established:

- developed information technology for classification of fractal ordered data sets, which is based on a combination of decision tree methods and neural networks, takes into account self-similar and multifractal properties of ordered data, which increases the accuracy of classification;

- a method of estimating the Hurst exponent according to the values of the time series is proposed, which is based on the application of classification methods of machine learning, which allows to significantly reduce the confidence interval of evaluation;

- applied recurrence characteristics of time series, as features for the classification of time series based on their fractal properties by machine learning methods, allowed to increase the accuracy of classification.

Further development was received by:

- a method of generating fractal realizations based on a stochastic multiplicative binomial cascade, which, unlike the existing ones, uses analytical determination of asymmetric beta distribution parameters, which allows to generate ordered data arrays with given multifractal properties;

- methods of binary classification of time series with fractal properties, based on the use of decision trees and neural networks, which, unlike the existing ones, use fractal and recurrence characteristics as features, which allows to increase the accuracy of classification.

The practical significance of the obtained results is that as a result of the thesis research a set of methods, algorithms and programs was created, which allow to solve the problem of classification of fractal ordered data sets. All theoretical developments of the thesis are brought by the author to concrete methods and are put in a basis of functioning of information technology of classification of the ordered data sets with fractal properties by methods of machine learning.

The proposed method of estimating the Hurst exponent, which is implemented in software, allows to improve the accuracy of its estimation according to ordered data, namely: reduces the confidence interval of the estimation several times.

The analysis of existing methods of classification of ordered data sets showed that, despite the fact that the use of machine learning in fractal data analysis is constantly growing, to date, no information technology has been developed for the classification of fractal data sets, including time series.

The thesis reviews the existing methods of generating fractal and multifractal ordered data sets. The necessity of improvement of such methods is proved, and the approach to generation of a stochastic multiplicative binomial cascade on the basis of the decision of a problem of nonlinear optimization is offered. Solving this problem allows the use of an asymmetric beta distribution to generate a stochastic multiplicative binomial cascade, which allows us to cover a wide range of multifractal properties. The use of an advanced method for generating a stochastic multiplicative binomial cascade makes it possible to generate more complex fractal realizations of multifractal Brownian motion.

Numerical experiments on the classification of different types of ordered data sets, which were divided into classes according to their fractal properties, were performed. The following classification methods were used: decision trees, bagging, random forest, multilayer perceptron, long-term short-term memory networks. Studies have shown that the range of multifractal and self-similar properties plays an important role in the choice of classifier, and accordingly on the accuracy of classification. It was proved that the use of statistical and fractal characteristics of multifractal ordered data sets improves the accuracy of ODS classification compared to the use of ODS values as features in the classification.

Recurrence plots are considered and the use of their characteristics as features in the training of classification models is proposed. Studies have shown that the characteristics of recurrence plots make a significant contribution to the accuracy of classification, and allow to classify ODS with monofractal properties.

The information technology of ODS classification is developed, which allows to choose the optimal method of machine learning for the classification of the studied ordered data set.

The use of the developed information technology for determination of the Hurst exponent of ODS is offered. The effectiveness of using machine learning methods to determine the Hurst exponent and its confidence interval has been experimentally proven, and it has allowed to reduce the confidence interval of the Hurst exponent for fractal and multifractal ODS more than 2 times.

The results of the application of the proposed information technology to detect DDoS attacks in a computer network are presented. IT increases the number of detected attacks, which allows you to use the proposed IT as an additional filter in detecting DDoS attacks.

The models, methods and information technology of fractal data classification proposed in the thesis were introduced into the activity of enterprises of different branches of economy:

– for the enterprise PrJSC "Farlep-Invest" (Kharkiv) the developed information technology for reduction of negative influence of cyber-attacks on information systems of the enterprise is introduced;

– for the enterprise CB "PrivatBank" (Kharkiv) the developed information technology for detection of intrusions into information systems of the enterprise is introduced.

The results of the dissertation research were introduced in the educational process of Kharkiv National University of Radio Electronics at the Department of Infocommunication Engineering in the teaching of the discipline "Artificial Intelligence (Artificial Intelligence)".

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

List of applicant publications

Scientific papers, in which the main scientific results of the thesis are published:

1. Bulakh V.A. Modelirovaniye mul'tiplikativnykh stokhasticheskikh kaskadov na osnove sluchaynoy velichiny s beta raspredeleniyem. *International journal information content & processing*. Bulgaria, 2017. Vol. 4, No 2. P. 159–170.

2. Kirichenko L.O., Bulakh V.A., Radivilova T.A. Klassifikatsiya mul'tifraktal'nykh stokhasticheskikh vremennykh ryadov s ispol'zovaniyem meta-algoritmov na osnove derev'yev resheniy. *Sistemní tekhnologíí*. Dnípro, 2018. № 3(116). S. 22–27.

3. Kirichenko L., Bulakh V., Radivilova T., Chernyak V. Analiz vzaimozavisimosti vremennykh ryadov bitkoina i aktivnosti soobshchestv v sotsial'nykh setyakh. *International journal information technologies & knowledge*. Bulgaria, 2018. Vol. 12, No 1. P. 43–55.

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

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.

6. Radivilova T., Kirichenko L., Bulakh V. Obnaruzheniye DDoS-atak metodami mashinnogo obucheniya na osnove fraktal'nykh svoystv. *Security in cyberspace, the social internet space in context values and hazards*. Slupsk-Charkow, Poland, 2019. P. 299–315.

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.

8. Kirichenko L.O., Bulakh V.A., Tavalbekh M.F., Zinchenko P.P. Informatsiyna tekhnolohiya klasyfikatsiyi fraktal′nykh chasovykh ryadiv. *Systemni tekhnolohii*. 2020. № 3 (128). S. 115–126.

9. Bulakh V.A. Informatsiyna tekhnolohiya klasyfikatsiyi uporyadkovanykh masyviv danykh z fraktal′nymy vlastyvostyamy. *Radioelektronika ta informatyka*. Kharkiv, 2020. № 2 (89). S. 66–72.

10. Kirichenko L.O., Bulakh V.A., Radivilova T.A., Tavalbekh M.F., Zinchenko P.P. Balansuvannya samopodibnoho trafiku v merezhnykh systemakh vyyavlennya vtorhnen′. *Kiberbezpeka: osvita, nauka, tekhnika*. Kyyiv, 2020. №3 (7). S. 17–32. URL: https://csecurity.kubg.edu.ua/index.php/journal/article/view/130.

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.

Materials and works of scientific conferences:

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.

13. Kirichenko L.O., Radivilova T.A., Bulakh V.A. Mul′tyfraktal′nyy analiz chasovykh ryadiv v sotsial′nykh merezhakh. *Praktychne zastosuvannya neliniynykh dynamichnykh system v infokomunikatsiyakh*: materialy 6 mizhnarodnoyi naukovo-praktychnoyi konferentsiyi (1 mizhnarodnyy sympozium). Chernivtsi, 2017. S. 19–20.

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.

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.

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.

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.

18. Kirichenko L, Bulakh V., Radivilova T., Kobitska Y. Fractal analysis of DDoS attack realizations. *Fizyko-tekhnichni problemy peredavannya, obrobky ta zberihannya informatsiyi v infokomunikatsiynykh systemakh*: Materialy 7-oyi mizhnarodnoyi naukovo-praktychnoyi konferentsiyi. 8-10 lystopada 2018 r. Chernivtsi, 2018. S. 13.

19. Radivilova T., Bulakh V., Kirichenko L. The method to control buffering and bandwidth network by monitoring of fractal traffic and security parameters. *Fizyko-tekhnichni problemy peredavannya, obrobky ta zberihannya informatsiyi v infokomunikatsiynykh systemakh*: materialy 7-oyi mizhnarodnoyi naukovo-praktychnoyi konferentsiyi. Chernivtsi, 2018. S.97.

20. Kyrychenko L.O., Bulakh V.A., Radyvylova T.A. Sravnytel′nyy analyz klassyfykatsyy mul′tyfraktal′nykh vremennykh ryadov. *Informatsiyni tekhnolohiyi v metalurhiyi ta mashynobuduvanni*: materialy mizhnarodnoyi naukovo-tekhnichnoyi konferentsiyi. 27 – 29 bereznya 2018 r. Dnipro, 2018. S.126.

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.

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

23. Kyrychenko L.O., Bulakh V.A. Klassyfykatsyya vremennykh ryadov na osnove fraktal′nykh svoystv. *Mizhnarodnyy naukovyy sympozium Intelektual'ni rishennya*: materialy V-oyi mizhnarodnoyi naukovo-praktychnoyi konferentsiyi. 15–20 kvitnya 2019r. Uzhhorod, 2019. S. 220.

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.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ 4

ВСТУП 5

Розділ 1. Фрактальні та мультифрактальні процеси як потік вхідної інформації для інформаційної технології класифікації упорядкованих масивів даних 12

1.1 Фрактальні часові ряди та упорядковані масиви даних 12

1.1.1 Самоподібний стохастичний процес з неперервним часом 12

1.1.2 Самоподібний стохастичний процес з дискретним часом 14

1.1.3 Довгострокова залежність та важкі хвости розподілів 15

1.2 Параметр Херста та методи його обчислення по експериментальним даним 17

1.3 Мультифрактальні процеси 18

1.4 Методи мультифрактального аналізу 25

1.5 Методи машинного навчання 28

1.6 Вимоги до інформаційної технології 37

1.7 Постановка задачі дослідження 39

ВИСНОВКИ ПО ПЕРШОМУ РОЗДІЛУ 40

Розділ 2. Використання моделювання фрактальних упорядкованих масивів даних як метод доповнення даних для інформаційної технології в умовах недостатньої інформації 42

2.1 Біноміальний стохастичний мультиплікативний каскад 43

2.2 Моделювання біноміальних стохастичних мультиплікативних каскадів, на основі оцінок мультифрактальних характеристик часових рядів 48

2.3 Експоненціальне перетворення фрактального гаусівського шуму 51

2.4 Мультифрактальний броунівський рух. Побудова реалізацій фрактального броунівського руху в мультифрактальному часі на основі каскадної моделі 56

ВИСНОВКИ ПО ДРУГОМУ РОЗДІЛУ 58

Розділ 3. Методи машинного навчання як інструмент для функціонування інформаційної технології класифікації упорядкованих масивів даних 60

3.1. Розбиття на класи 61

3.2. Класифікатори 62

3.3. Характеристики якості класифікації для порівняння 64

3.4. Ознаки 65
3.4.1. Ознаки – значення часового ряду 65
3.4.2. Ознаки – статистичні та мультифрактальні характеристики 71
3.4.3. Ознаки – фрактальні та рекурентні характеристики 74
3.5 Оцінювання показника Херста методами машинного навчання 85
ВИСНОВКИ ПО ТРЕТЬОМУ РОЗДІЛУ 87
Розділ 4. Інформаційна технологія класифікації фрактальних упорядкованих масивів даних 89
4.1. Розробка інформаційної технології класифікації фрактальних упорядкованих масивів даних 89
4.2. Застосування запропонованої ІТ на прикладі детектування DDoS-атак 94
ВИСНОВКИ ПО ЧЕТВЕРТОМУ РОЗДІЛУ 101
ВИСНОВКИ 102
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 104
Додаток А. Документи впровадження основних результатів дисертаційної роботи 119
Додаток Б. Список публікацій здобувача 123

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

УМД – упорядкований масив даних

МФДФВ – мультифрактальний детрендований флуктуаційний аналіз

ФБР – фрактальний броунівський рух

ФГШ – фрактальний гаусівський шум

ЕПФГШ – експоненціальне перетворення фрактальоного гаусівського шуму

МФБР – мультифрактальних броунівський рух

СМБК – стохастичний мультиплікативний біноміальний каскад

РД – рекурентні діаграми

ІТ – інформаційна технологія

ДР – дерева регресії, регресійні дерева підтримки прийняття рішень

ДК – дерева класифікації, класифікуючі дерева підтримки прийняття рішень

Б[ДР] – Беггінг з використанням регресійних дерев підтримки прийняття рішень

ВЛ[ДР] – випадковий ліс з використанням регресійних дерев підтримки прийняття рішень

НМ – нейронні мережі

БП – багатошаровий персептрон

DDoS – Distributed Denial of Service

LSTM – Long short-term memory

ВСТУП

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:

1. ІТ має збирати вхідні дані для аналізу та в on-line режимі, їх оброблювати та зберігати до бази даних
2. В умовах недостатньої кількості даних ІТ має використовувати методи моделювання для генерації даних з подібними фрактальними властивостями до аналізованих процесів

3. Використовуючи методи машинного навчання, згенеровані та отримані реальні дані, ІТ повинна навчитися оптимально виявляти (класифікувати) різні стани досліджуваного об’єкту

4. Використовуючи навчену модель класифікувати вхідні дані та інформувати зовнішню систему та оператора ІТ при виявленні особливостей у вхідних даних

5. Зберігати до бази даних всі результати роботи ІТ

6. Зі збільшенням обсягів реальних даних проводити перенавчання моделі класифікації

Зв'язок роботи з науковими програмами, планами, темами. Результати дисертаційної роботи використані в ході виконання науково-дослідних робіт Харківського національного університету радіоелектроніки (ХНУРЕ) у межах договору за замовленням Міністерства освіти і науки України НДР (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 в роботі використовувався математичний апарат методів машинного навчання; методи математичної статистики – для статистичного аналізу упорядкованих множин даних, аналізу точності класифікації методами машинного навчання; методи теорії хаосу – для обчислення фрактальних та рекурентних характеристик; методи математичної статистики – для аналізу якості моделей машинного навчання.

Наукова новизна отриманих результатів полягає в наступному

вперше:

розроблено інформаційну технологію класифікації фрактальних упорядкованих масивів даних, яка заснована на поєднанні методів дерев рішень та нейронних мереж, враховує самодібні та мультифрактальні властивості упорядкованих даних і дозволяє збільшити точність класифікації;

запропоновано метод оцінювання показника Херста за значеннями часового ряду, який засновано на застосуванні класифікаційних методів машинного навчання, що дозволяє суттєво збільшити точність оцінювання;

застосовано рекурентні характеристики часових рядів як ознаки для класифікації часових рядів на основі їхніх фрактальних властивостей методами машинного навчання, що дозволило 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. Всі 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 кібер-атак 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 кібер-атак в інфокомунікаційних даних, уточнення діагнозу за даними записів ЕЕГ та ЕКГ, класифікації сейсмічних подій за сейсмограмами тощо.

Особистий внесок здобувача. Дисертаційна робота є результатом самостійної роботи автора. Наукові положення, висновки і рекомендації, які становлять суть дисертаційної роботи, були сформульовані, розроблені і досліджені самостійно. За результатами дисертаційної роботи було опубліковано

24 праці. В роботах, написаних у співавторстві, здобувачеві належать: [2] – розроблений метод класифікації фрактальних часових рядів, [3] – проведено мультифрактальних аналіз часових рядів бітколїна та активності в соціальних мережах, [4] – проведено дослідження узагальненого підходу до оцінки показника Херста за часовими рядами, [5] – запропоновано методи класифікації часових рядів на основі фрактальних властивостей, [6] – досліджено мультифрактальні властивості та запропоновано методи класифікації часових рядів, [7] – проведено порівняльний аналіз методів класифікації нейронними мережами та деревами прийняття рішень, [8] – розроблена та запропонована інформаційна технологія класифікації фрактальних часових рядів, [10] – запропоновано використовувати методи машинного навчання для виявлення факту вторгнення в комп'ютерну мережу, [11] – проведено дослідження методів виявлення вторгнень в комп'ютерні мережі, [12] – досліджено взаємозв'язок рядів біткоїна та активності соціальних мереж методами фрактального аналізу, [13] – проведено мультифрактальних аналіз та досліджено взаємозв'язок рядів біткоїна та активності соціальних мереж, [14] – проведено дослідження методів виявлення вторгнень в комп'ютерні мережі, [15] – проведено порівняльний аналіз методів класифікації нейронними мережами та деревами прийняття рішень, [16] – проведено порівняльний аналіз класифікації часових рядів заснованої на фрактальних характеристиках, [17] – проведено порівняльний аналіз класифікації часових рядів заснованої на рекурентних характеристиках, [18] – проведено фрактальний аналіз DDoS-атак, [19] – досліджено мультифрактальні властивості та запропоновано методи класифікації часових рядів, [20] – запропоновано використання фрактальних, рекурентних характеристик часових рядів як ознак при машинному навчанню, [21] – проведено фрактальний аналіз DDoS-атак, запропоновано метод виявлення DDoS-атак заснований на машинному навчанні, [22] – запропоновано використання фрактальних, рекурентних характеристик часових рядів як ознак при машинному навчанню, [23] – проведено порівняльний аналіз класифікації фрактальних процесів, [24] – промодельовано мультифрактальні часові ряди.

Апробація результатів дисертації. Основні результати дисертаційної роботи доповідались та обговорювались на таких конференціях:

- 4th international scientific-practical conference «Problems of infocommunications. Science and technology (PIC S&T)»(Kharkiv, 2017);
- 9-й IEEE International conference on dependable systems, services and technologies (DESSERT)(Kyiv, 2018);
- Proceedings of the 14th international conference on ICT in education, research and industrial applications. Integration, harmonization and knowledge transfer (Kyiv, 2018);
- IEEE Second international conference on data stream mining & processing (DSMP) (Lviv, 2018);
- International scientific-practical conference problems of infocommunications. Science and technology (PIC S&T) (Kharkiv, 2018);
- 10th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (2019);
- 2019 IEEE 8th international conference on advanced optoelectronics and lasers(2019);
- third international workshop on computer modeling and intelligent systems(Zaporizhzhia, 2020);
- 6-а Міжнародна науково-практична конференція (1-й міжнародний симпозіум). Практичне застосування нелінійних динамічних систем в інфокомунікаціях (Чернігів, 2017);
- 7-а Міжнародна науково-практична конференція фізико-технічні проблеми передавання, обробки та зберігання інформації в інфокомунікаційних системах (Чернівці, 2018);
- Міжнародна науково-технічна конференція інформаційні технології в металургії та машинобудуванні (Дніпро, 2018);
- 5-а міжнародна науково-практична конференція. Міжнародний науковий симпозіум «Інтелектуальні рішення» (2019).

- 5-а міжнародна науково-практична конференція. Міжнародний науковий симпозіум «Інтелектуальні рішення» Ужгород (2019).

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 зі вступу, чотирьох розділів, висновків, списку використаних джерел з 134 найменувань та 2 додатка; містить 20 таблиць і 27 рисунків та має загальний обсяг 127 сторінок.

РОЗДІЛ 1
ФРАКТАЛЬНІ ТА МУЛЬТИФРАКТАЛЬНІ ПРОЦЕСИ ЯК ПОТІК ВХІДНОЇ ІНФОРМАЦІЇ ДЛЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ КЛАСИФІКАЦІЇ УПОРЯДКОВАНИХ МАСИВІВ ДАНИХ

Оскільки ми пропонуємо інформаційну технологію для класифікації фрактальних процесів методами машинного навчання, необхідно навести поняття фрактальних та мультифрактальних процесів та їх властивості. Також необхідно визначити методи машинного навчання котрі будемо використовувати для побудови класифікаторів.

1.1 Фрактальні часові ряди та упорядковані масиви даних

Основні поняття і визначення самоподібних випадкових процесів були приведені в роботах [71-74, 76, 96, 111-116, 121].

1.1.1 Самоподібний стохастичний процес з неперервним часом

Стохастичний процес $X(t), t \in \mathbb{R}$ з неперервною змінною часу називається самоподібним, у вузькому сенсі, з параметром $H, 0 < H < 1$, якщо для будь-якого дійсного значення $a > 0$ скінченномірні розподіли для ідентичні скінченномірним розподілам $a^{-H} X(at), t \in \mathbb{R}$ тобто, якщо для будь-яких $k \geq 1, t_1, t_2, ..., t_k \in \mathbb{R}$ та будь-яких $a > 0$

$$Law\{X(t_1), ..., X(t_k)\} = Law\{a^{-H} X(at_1), ..., a^{-H} X(at_k)\}. \quad (1.1)$$

Стислий вигляд рівняння (1.1)

$$Law\{X(t), t \in \mathbb{R}\} = Law\{a^{-H} X(at), t \in \mathbb{R}\}. \quad (1.2)$$

Позначення $Law\{\cdot\}$ означає скінченномірні закони розподілу випадкового процесу. Параметр H, називається параметром Херста та представляє собою міру самоподібності стохастичного процесу.

Стохастичний процес $X(t)$ являється стохастично самоподібним або самоподібним в широкому сенсі, якщо процес $a^{-H}X(at)$ володіє такими ж статистичними характеристиками другого порядку (математичним очікуванням $M[X(t)]$, дисперсією $D[X(t)]$ та автокореляційною функцією $r_x(t,s)$), що і $X(t)$:

$$M[X(t)] = \frac{M[X(at)]}{a^H};$$

$$D[X(t)] = \frac{D[X(at)]}{a^{2H}};$$

$$r_x(t,s) = \frac{r_x(at,as)}{a^{2H}}.$$

Можна показати, поклавши $a = \frac{1}{t}$, що для самоподібного процесу виконується наступна рівність

$$Low\{X(t)\} = Low\left\{\left(\frac{1}{t}\right)^{-H} X(1)\right\} = Law\{t^H X(1)\}. \quad (1.3)$$

Враховуючи (1.3), початкові моменти самоподібного випадкового процесу можна виразити як

$$M\left[|X(t)|^q\right] = M\left[|t^H X(1)|^q\right] = t^{qH} M\left[|X(1)|^q\right] = C(q)\cdot t^{qH}, \quad (1.4)$$

де величина $C(q) = M\left[|X(1)|^q\right]$.

Важливим класом випадкових процесів, які ми можемо зустріти в різних областях природознавства та техніки, являються випадкові процеси, що мають стаціонарні прирощення. Цей процес $X(t), t \in \mathbb{R}$, називається самоподібним, якщо

$$Low\{X(t+\Delta t) - X(\Delta t)\} = Low\{X(t) - X(0)\},\ \forall t, \Delta t \in \mathbb{R}. \quad (1.5)$$

Такі процеси мають середнє значення, що дорівнює 0, дисперсія визначається як $M\left[X^2(t)\right]=\sigma^2|t|^{2H}$, а кореляційна функція залежить лише від моментів часу t та s

$$M\left[X(t),X(s)\right]=\frac{\sigma^2}{2}\left(|t|^{2H}+|s|^{2H}-|t-s|^{2H}\right). \tag{1.6}$$

1.1.2 Самоподібний стохастичний процес з дискретним часом

Розглянемо поняття самоподібності для процесу з дискретним часом [126, 134]. Нехай $X=\left(X_1, X_2,...\right)$ – відрізок стаціонарного в широкому сенсі випадкового процесу з дискретним часом $t\in\mathbb{N}=\{1, 2, 3,...\}$. Припустимо, що процес X має автокореляційну функцію наступного виду

$$r(k)\sim k^{-\beta}L_1(k),\ k\to\infty, \tag{1.7}$$

де $0<\beta<1$ і L_1 – функція повільно змінюється на нескінченності, тобто $\lim\limits_{t\to\infty}\frac{L_1(tx)}{L_1(t)}=1$ для всіх $x>0$.

Позначимо через $X^{(m)}=\left\{X_1^{(m)},X_2^{(m)},...\right\}$ – усереднений по блокам довжини m процес X, компоненти якого визначаються рівністю

$$X^{(m)}=\frac{1}{m}\left(X_{tm-m+1}+...+X_{tm}\right),\ m,t\in\mathbb{N}. \tag{1.8}$$

Такий ряд називається агрегованим. Позначимо через $r_m(k)$ і D_m коефіцієнти кореляції та дисперсію процесу $X^{(m)}$ відповідно. Процес X називається строго самоподібним, в широкому сенсі, з параметром $H=1-\frac{\beta}{2},0<\beta<1,$ якщо

$$r_m(k)=r(k),k\in\mathbb{Z}_+,m\in\{2, 3,...\}.$$

Тобто, процес не змінює свій коефіцієнт кореляції після усереднення по блокам довжиною m. Іншими словами, X є самоподібним в широкому сенсі,

якщо агрегований процес $X^{(m)}$ не відрізняється від початого процесу X, у відношенні статистичних характеристик другого порядку.

Процес X називається асимптотично само подібним, в широкому сенсі, з параметром $H = 1 - \frac{\beta}{2}, 0 < \beta < 1,$ якщо

$$\lim_{m \to \infty} r_m(k) = r(k), k \in \mathbb{N}.$$

Процес X називається самоподібним, у вузькому сенсі, з параметром $H = 1 - \frac{\beta}{2}, 0 < \beta < 1,$ якщо справедлива рівність

$$Low\left\{m^{1-H} X^{(m)}\right\} = Low\{X\}, m \in \mathbb{N}. \quad (1.9)$$

1.1.3 Довгострокова залежність та важкі хвости розподілів

Поняття повільно спадної залежності має ключове значення в теорії самоподібних процесів та описує довгострокову пам'ять процесу: майбутнє процесу визначається його минулим, при цьому зі спадаючим ступенем впливу по мірі того, як минуле віддалене від теперішнього [5, 10, 15, 115]. Таким чином, процес з довгостроковою пам'яттю якби «забуває» своє відносно далеке минуле по мірі руху часу в майбутнє.

Процес X володіє повільно спадаючою залежністю, якщо виконується співвідношення (1.7). Таким чином, процеси з повільно спадаючою залежністю характеризуються автокореляційною функцією, котра спадає гіперболічно – по ступеневому закону – при збільшенні часової затримки. Можна показати, що з (1.7) слідує несумованість автокореляційної функції, тобто $\sum_k r(k) = \infty$. На відміну від процесів з повільно спадаючою залежністю, процеси з швидко спадаючою залежністю володіють експоненціально спадаючою автокореляційною функцією виду

$$r(k) \sim \rho^k, k \to \infty, 0 < \rho < 1,$$

і, як наслідок, сумованістю автокореляційної функції $0 < \sum_k r(k) < \infty$.

Дискретні самоподібні процеси володіють повільно спадаючою автокореляційною функцією агрегованого процесу $X^{(m)}$ при $m \to \infty$, на відміну від поширених стохастичних моделей для яких виконується

$$r_m(k) \to 0,\ m \to \infty,\ k \in \mathbb{N}.$$

Завдання автокореляційної функції у вигляді (1.7) рівнозначно визначенню процесу з наступною дисперсією

$$\sigma^2 \sim am^{-\beta},\ m \to \infty,$$

де a - кінцева додатна константа котра не залежить від $m, 0 < \beta < 1.$

Тобто, довгострокова залежність має на увазі властивість самоподібності, в широкому сенсі, зі значенням показника $H > 0.5$ і навпаки [17, 25].

Більшість процесів, котрі мають довгострокову залежність, володіють важкими хвостами одновимірної функції розподілу ймовірностей [112]. Випадкова величина X має розподілення з важкими хвостами, якщо

$$P[X > x] \sim c \cdot X^{-a},\ x \to \infty, \qquad (1.10)$$

де величина a, $0 < a < 2,$ називається параметром форми, c - деяка додатна константа.

На відміну від розподілу з легкими хвостами, такими як експоненціальне або гаусовське, котрі мають експоненціальне спадання хвоста, випадкова величина з важким хвостом володіє хвостом, котрий спадає по степеневому (гіперболічному) закону. При $0 < a < 2$ випадкові величини мають нескінченну дисперсію, а при $0 < a < 1$ ще і мають нескінченне середнє. Основні властивості випадкової величини, розподіленої з важким хвостом, складається в тому, що вона проявляє високу змінність. Іншими словами, її вибірка представляє собою в більшості відносно невеликі значення, проте також містить і достатня кількість дуже великих значень.

1.2 Параметр Херста та методи його обчислення по експериментальним даним

Параметр Херста H – званий параметром самоподібності – знаходиться в діапазоні $0 < H < 1$ та представляє собою ключову міру самоподібності та тривалості довгострокової залежності стохастичного процесу [113, 114, 120].

У випадку $0{,}5 < H < 1$ говорять про персистентну (підтримуючу) поведінку процесу або про те, що процес володіє довгостроковою пам'яттю. Іншими словами, якщо протягом деякого часу в минулому спостерігались додатні збільшення процесу, тобто відбувалось збільшення, і надалі в середньому будуть відбуватися збільшення. Іншими словами, ймовірність того, що процес на $i+1$ кроці відхиляється від середнього в тому ж напрямку, що і на кроці i настільки велика, наскільки параметр H близький до 1. Таким чином, персистентні стохастичні процеси виявляють чітко виражені тенденції змінення при відносно малому шумі.

При $H = 0{,}5$ відхилення процесу від середнього являється дійсно випадковим та не залежить від попередніх значень, що відповідає випадку звичайного броунівського руху.

Безпосередні розрахунки ступеня гіперболічного спадання кореляційної функції часового ряду $K(\tau)$ ускладнені у зв'язку з тим, що, починаючи з деякого значення затримки, τ значення $K(\tau)$ стають порівняними з помилками обчислень, виникаючими при аналізі рядів кінцевої довжини. Таким чином, в рамках класичного кореляційного аналізу стає неможливим відслідковувати закономірності спаду кореляцій при довгостроковій залежності.

Метод детрендованого флуктуаційного аналізу (ДФА) був запропонований для аналізу довготривалих кореляцій в структурі сердечного ритму[52]. У теперішній час він являється одним з найбільш широко використовуваним методом дослідження довготривалих кореляцій у нестаціонарних процесах [52, 53]. Зокрема, в дослідженнях структури ДНК [52], в дослідженні динаміки біологічних систем по експериментальним даним, в останні роки ДФА активно

використовується у фрактальному аналізі фінансових часових рядів для прогнозування корисних ситуацій [29, 84].

В методі ДФА для вихідного часового ряду $x(t)$ будується кумулятивний ряд $y(t)=\sum_{i=1}^{t}x(i)$, котрий розбивається на N ділянок довжиною τ, та для кожного сегмента $y(t)$ обчислюється флуктуаційна функція

$$F^2(\tau)=\frac{1}{\tau}\sum_{t=1}^{\tau}\left(y(t)-Y_m(t)\right)^2, \qquad (1.11)$$

де $Y_m(t)$ - локальний m- поліноміальний тренд в рамках даного сегменту.

Функція $F(\tau)$, усереднення у всьому ряду $y(t)$, володіє скеійлінговою залежністю від довжини ряду: $F(\tau)\propto\tau^H$. Графік залежності $\log F(\tau)$ від $\log(\tau)$ в деякому діапазоні значень буде являти собою пряму лінію, апроксимована методом найменших квадратів. Оцінка показника H обчислюється як тангенс кута нахилу прямої залежності $\log F(\tau)$ вид $\log(\tau)$.

1.3 Мультифрактальні процеси

Мультифрактальні міри буди введені Б.Мандельбротом [72, 73] для опису розподілення турбулентної дисипації. Із тих пір вони були знайдені у множині природних та технічних процесів та використовуються для моделювання та дослідження турбулентних процесів, телекомунікаційних потоків, фінансових ринків, медичних даних, геофізичних процесів.

Мультифрактальність – це концепція, котра, з деякими незначними зміненнями може бути в рівнім ступені добре застосована до функцій і мір детермінованим та стохастичним. В описі основних понять та властивостей мультифрактальних процесів існує декілька підходів: глобальний, при якому основні визначення та властивості виводять через властивості моментів випадкових процесів [11, 96]; локальний – описує локальні масштабні та сингулярні властивості фрактальних реалізацій і функцій та заснований на властивостях локальних експонент Гельдера [3, 4, 71]; та підхід, заснований на

властивостях фрактальних і мультифрактальних мір [96, 104].У теперішньому часі з’явилась низка монографій, де зроблені спроби поєднати всі підходи в одне ціле [3, 96].

Закони розподілення та моментні характеристики мультифрактальних процесів

На відміну від самоподібних процесів (1.2), мультифрактальні процеси володіють більш різноманітними скейлінговими характеристиками

$$\mathrm{Law}\left\{X\left(at\right)\right\}=\mathrm{Law}\left\{M\left(a\right)\cdot X\left(t\right)\right\},\ a>0, \tag{1.12}$$

де $M\left(a\right)$ - незалежна від $X\left(a\right)$ випадкова функція.

У випадку самоподібного процесу $M\left(a\right)=a^{H}$. Для мультифрактальних процесів узагальнений показник Херста $H\left(a\right)=\log_{a}M\left(a\right)$ є випадковою функцією аргументу a. Співвідношення (1.12) можливо переформулювати наступним чином

$$\mathrm{Law}\left\{X\left(at\right)\right\}=\mathrm{Law}\left\{a^{H(a)}X\left(t\right)\right\}. \tag{1.13}$$

Властивість мультифрактальності (1.13) встановлює достатньо жорсткі обмеження на закони розподілу випадкових величин $M\left(a\right)$ та $X\left(t\right)$. Для випадкової слеклінгової функції $M\left(a\right)$ має виконуватися наступна умова

$$\mathrm{Law}\left\{M\left(ab\right)\right\}=\mathrm{Law}\left\{M_{1}\left(a\right)\cdot M_{2}\left(b\right)\right\}, \tag{1.14}$$

де $M_{1}\left(a\right)$ та $M_{2}\left(a\right)$ незалежні та мають такий самий закон розподілення що і $M\left(a\right)$.

Для мультифрактального процесу $X\left(t\right)$ має слідувати, що якщо виконується умова $\frac{a_{2}}{a_{1}}=\frac{a_{3}}{a_{2}}$, то співвідношення (1.13) тягне за собою наступне

$$\mathrm{Law}\left\{\frac{X\left(a_{2}t\right)}{X\left(a_{1}t\right)}\right\}=\mathrm{Law}\left\{\frac{X\left(a_{3}t\right)}{X\left(a_{2}t\right)}\right\}.$$

З цих властивостей витікає визначальна властивість мультифрактальних процесів [12, 71]; процес $X(t)$ є мультифрактальним, якщо виконується наступне співвідношення

$$M\left[\left|X(t)\right|^{q}\right]=c(q)\cdot t^{\tau(q)+1},\ \forall \mathrm{t}\in \mathrm{T},\ \forall \mathrm{q}\in \mathbb{Q}, \tag{1.15}$$

де $c(q)$ та $\tau(q)$ - деякі детерміновані функції. Величини T та $\mathbb{Q}$ є інтервалами ненульової довжини на дійсній вісі, та $0\in T$, $[0,1]\subset \mathbb{Q}$.

Мультифрактальні процеси, таким чином, володіють глобальним скейлінгом, в тому сенсі, що його моменти задовольняють скейлінговому співвідношенню (1.15). Функція $\tau(q)$ називається скейлінговою експонентою мультифрактального процесу. З (1.15) витікають деякі властивості функції $\tau(q)$: $\tau(0)=-1;\ \tau(1)=0$, $\tau(q)$ - випукла функція.

Узагальнений показник Херста $h(q)$ зв'язаний з функцією $\tau(q)$ співвідношенням [52, 53]

$$\tau(q)=qh(q)-1. \tag{1.16}$$

Самоподібні процеси є підкласом мультифрактальних. Можна показати, поклавши в (1.2) $a=\frac{1}{t}$, що для самоподібних процесів виконується наступна рівність

$$\mathrm{Law}\left\{X(t)\right\}=\mathrm{Law}\left\{\left(\frac{1}{t}\right)^{H}X(1)\right\}=\mathrm{Law}\left\{t^{H}X(1)\right\}. \tag{1.17}$$

Враховуючи (1.17), моменти самоподібного випадкового процесу, можна виразити як:

$$M\left[\left|X(t)\right|^{q}\right]=M\left[\left|t^{H}X(1)\right|^{q}\right]=t^{qH}M\left[\left|X(1)\right|^{q}\right]=C(q)\cdot t^{qH}, \tag{1.18}$$

де величина $C(q)=M\left[\left|X(1)\right|^{q}\right]$.

Для самоподібних процесів скейлінгова експонента лінійна: $\tau(q) = Hq - 1$ та значення $\frac{\tau + 1}{q}$ при $q = 2$ співпадає зі значенням ступені самоподібності H.

Локальні показники Гельдера для мультифрактальних процесів

Цей підхід заснований на локальних показниках Гельдера - фундаментальна концепція в дійсному аналізі, котрий описує масштабні властивості реалізації в будь-який момент часу t_0.

Для реалізації самоподібного процесу можна записати наступний вираз

$$\left|g(t_0 + at) - g(t_0)\right| \sim a^H \left|g(t_0 + t) - g(t_0)\right|. \quad (1.19)$$

Показник ступеня H в формулі (1.19) є показником Херста та в даному випадку характеризує, наскільки нерегулярна функція $g(t)$ в точці t_0. При $H < 1$ ця функція є не диференційною та, чим менше H, тим більше сингулярна функція $g(t)$. У межах мультифрактального аналізу реалізації випадкового процесу розглядаються не з точки зору складної геометрії, а як функції з локальним нерегулярним поводженням. Враховуючи те, що сингулярності функції можуть відрізнятися в різних точках, їх опис в загальному випадку не може відбуватися за допомогою одного параметру H. У зв'язку з цим використовують поняття локальних експонент Гельдера $\alpha(t_0)$

$$\left|g(t_0 + \Delta t) - g(t_0)\right| \sim \Delta t^{\alpha(t_0)}, \quad (1.20)$$

Із (1.20) випливає, що показник Гельдера для неперервної функції $g(t)$ може бути визначений наступним чином:

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{\ln L(t, \Delta t)}{\ln(\Delta t)}, \quad (1.21)$$

де $L(t, \Delta t) = \left|g(t_0 + \Delta t) - g(t_0)\right|$.

У контексті реалізацій випадкових процесів визначення (1.21) передбачає метод оцінювання ймовірностей того, що випадково вибрана точка на відрізку

$[0,T]$ буде мати заданий показник Гельдера. Коротко опишемо процедуру наближеного визначення показника Гельдера: інтервал $[0,T]$ ітераційно ділиться на k-тому кроці на b^k під інтервалів; на кожному кроці для значення t_i обчислюються кінцеві величини $L\left(t_i, b^{-k}\cdot T\right)$ та приближено обчислюється показник Гельдера

$$\alpha_k\left(t_i\right)=\frac{\ln L\left(t_i,\mathrm{b}^{-k}\cdot T\right)}{\ln\left(b^k\right)}.$$

Розділимо ввесь діапазон значень α на малі непересічні інтервали $\left(\overline{\alpha_j},\overline{\alpha_j}+\Delta\alpha\right]$ та позначимо через $N_k\left(\overline{\alpha_j}\right)$ кількість значень $\alpha_k\left(t_i\right)$, що потрапили в нього. При $k\to\infty$ співвідношення $\frac{N_k\left(\overline{\alpha_j}\right)}{b^k}$ збігається до ймовірності того, що випадково вибрана на відрізку $[0,T]$ точка t буде мати показник Гельдера α.

Функція мультифрактального спектру $f(\alpha)$ визначається наступним чином

$$f(\alpha)=\lim_{k\to\infty}\frac{\ln N_k(\alpha)}{\ln \mathrm{b}^k}. \tag{1.22}$$

Якщо в функції $f(\alpha)$ границя існує та приймає додатне значення на інтервалі, відмінному від точки, тоді функція $g(t)$ володіє мультифрактальними властивостями.

Скейлінгова експонента $\tau(q)$ та функція мультифрактального спектра $f(\alpha)$ є випуклими вверх функціями та пов'язані між собою перетворенням Лежандра:

$$\begin{cases}\tau(q)=\min\limits_{\alpha}\left(\alpha q-f(\alpha)\right),\\ f(\alpha)=\min\limits_{q}\left(\alpha q-\tau(q)\right).\end{cases} \tag{1.23}$$

Прямий та зворотний зв'язок між мультифрактальним спектром та скейлінговою експонентою складає вміст мультифрактального формалізму. Мультифрактальний формалізм строго обґрунтована для великого класу мультифрактальних мір з гладкою строго ввігнутою функцією $\tau(q)$.

Характеристики мультифрактальних множин

Самоподібність фрактальних об'єктів полягає в збереженні структури об'єкта при зміненому масштабі. Розглянемо основні характеристики мультифрактальної множини [26, 104]: нехай в загальному випадку мультифрактальний об'єкт займає деяку обмежену область в d- мірному Евклідовому просторі визначає собою множину з $N \to \infty$ точок. Розіб'ємо всю область на кубічні комірки з ребром ε та об'ємом ε^d. Розглянемо узагальнену статистичну суму $Z(q,\varepsilon)$ яка характеризується показником ступені q, що може приймати будь-які значення в інтервалі $-\infty < q < +\infty$

$$Z(q,\varepsilon) = \sum_{i=1}^{N(\varepsilon)} P_i^q(\varepsilon), \tag{1.24}$$

де $P_i(\varepsilon) = \lim_{N\to\infty} \frac{n_i(\varepsilon)}{N}$;

$n_i(\varepsilon)$ - кількість точок, котрі попали в комірку за номером i;

$N(\varepsilon)$ - загальна кількість зайнятих комірок, котре залежить від розміру комірки ε.

Ймовірності p_i характеризують відносну зайнятість комірок. У загальному випадку мультифрактальна множина характеризується деякою нелінійною функцією $\tau(q)$, яка визначає поведінку статистичної суми $Z(q,\varepsilon)$ при $\varepsilon \to \infty$

$$Z(q,\varepsilon) \propto \varepsilon^{\tau(q)}. \tag{1.25}$$

Функція $\tau(q)$ зазвичай називається скейлінговою експонентою та визначається як

$$\tau(q) = \lim_{\varepsilon\to 0} \frac{\ln Z(q,\varepsilon)}{\ln \varepsilon}. \tag{1.26}$$

У випадку однорідної фрактальної множини з фрактальною розмірністю D у всіх зайнятих комірках знаходиться однакова кількість точок [8], тобто $p_i(\varepsilon) = p(\varepsilon) = \frac{1}{N(\varepsilon)}$, та узагальнена статистична сума приймає вигляд

$$Z(q,\varepsilon) = N^{1-q}(\varepsilon) = \varepsilon^{-D(1-q)},$$

а функція $\tau(q) = (q-1)D$ є лінійною. Якщо розподілення точок по коміркам неоднакове, то фрактальна множина є неоднорідним- мультифрактальним - та $\tau(q)$ є нелінійною функцією. При $q \to +\infty$ основний внесок в узагальнену статистичну суму вносять комірки, які містять найбільшу кількість частинок n_i, а, отже, характеризуються найбільшою ймовірністю заповнення p_i. Навпаки, при $q \to -\infty$ основний внесок сумарно надають найбільш розріджені комірки з малим значенням p_i. Таким чином, функція $\tau(q)$ показує, наскільки неоднорідним є досліджувана множина точок.

Функція мультифрактального спектра (спектра сингулярностей) $f(\alpha)$ в даному випадку вводиться наступним чином [47]: залежність ймовірностей від розміру комірок $p_i(\varepsilon)$ має ступеневий характер

$$p_i(\varepsilon) \propto \varepsilon^{\alpha_i}, \tag{1.27}$$

де α_i представляє собою деякий показник ступеню, різний для різник комірок (показник сингулярності). Для однорідного фракталу всі показники ступені α_i однакові та рівні фрактальній розмірності D.

Функція мультифрактального спектру $f(\alpha)$ характеризує розподілення ймовірностей різних значень α_i. Якщо величина $n(\alpha)d\alpha$ є ймовірністю того, що α_i знаходиться в інтервалі $(\alpha, \alpha + d\alpha)$, тобто представляє собою кількість комірок i, котрі володіють однаковою мірою $p_i(\varepsilon)$ з $\alpha_i \in (\alpha, \alpha + d\alpha)$, тоді

$$n(\alpha) \approx \varepsilon^{-f(\alpha)}. \tag{1.28}$$

Таким чином, функція $f(\alpha)$представляє собою фрактальну розмірність деякої однорідної фрактальної підмножини ξ_α з початкової множини ξ, котре характеризується однаковими ймовірностями заповнення комірок $p_i(\varepsilon) \approx \varepsilon^{\alpha}$.

Враховуючи вирази (1.24, 1.28), узагальнену статистичну суму $Z(q,\varepsilon)$ можна виразити через функцію мультифрактального спектру $f(\alpha)$ наступним чином

$$Z(q,\varepsilon)=\sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon)\approx\int d\alpha n(\alpha)\varepsilon^{q\alpha}\approx\int d\alpha\varepsilon^{q\alpha-f(\alpha)}.$$

Формально перехід від змінних $\{q,\tau(q)\}$ до змінних $\{\alpha, f(\alpha)\}$ може бути виконаний за допомогою наступного перетворення Лежандра:

$$\begin{cases}\alpha=\dfrac{d\tau}{dq},\\ f(\alpha)=q\dfrac{d\tau}{dq}-\tau,\end{cases}\quad\text{та}\quad\begin{cases}q=\dfrac{df}{d\alpha},\\ \tau(q)=\alpha\dfrac{df}{d\alpha}-f.\end{cases}\tag{1.29}$$

1.4 Методи мультифрактального аналізу

Метод детрендованого флуктуаційного аналізу

Даний метод використовується при оцінюванні параметру H для самоподібних часових рядів. У цьому випадку для вихідного часового ряду $x(t)$ будується кумулятивний ряд $y(t)=\sum_{i=1}^{t}x(t)$, який розбивається на N частин довжиною s. Для кожного сегменту $y(t)$ обчислюється флуктуаційна функція

$$F^2(s)=\frac{1}{s}\sum_{t=1}^{s}\left(y(t)-Y_m(t)\right)^2,\tag{1.30}$$

де $Y_m(t)$ локальний m- поліноміальний тренд в рамках даного сегменту.

Функція $F(s)$ усереднюється по всьому ряду $y(t)$. Такі обчислення повторюються для різноманітних розмірів сегментів, щоб отримати залежність $F(s)$ в широкому діапазоні значень параметра s. Для процесів з фрактальними властивостями зі збільшенням значення s функція $F(s)$ також зростає, та лінійна

залежність $\log F(s)$ від $\log s$,що свідчить про наявності властивостей масштабної інваріантності $F(s) \propto s^H$.

При дослідженні властивостей мультифрактальних процесів застосовується мультифрактальний флуктуаційний аналіз [53, 82]. При проведенні мультифрактального детрендованого флуктуаційного аналізу (МФДФА) використовується залежність флуктуаційної функції $F_q(s)$ від параметру q

$$F_q(s) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[F^2(s) \right]^{\frac{q}{2}} \right\}^{\frac{1}{q}}, \tag{1.31}$$

отриманою зведенням виразу (1.30) у ступінь q та наступним усередненням по всім сегментам. Оскільки при $q = 0$ вираз (1.31) містить невизначеність, то замість нього використовують наступний граничний вираз

$$F_q(\tau) = \mathrm{Exp}\left\{ \frac{1}{N} \sum_{i=1}^{N} \ln\left[F^2(\tau) \right] \right\}. \tag{1.32}$$

Змінюючи часову шкалу s при фіксованому показнику q, знаходимо залежність $F_q(s)$, представляючи її в подвійних логарифмічних координатах. Якщо досліджуваний ряд зводиться до мультифрактальної множини , котра проявляє довгострокові залежності, то флуктуаційна функція $F_q(s)$ представляється ступеневою залежністю

$$F_q(s) \propto s^{h(q)}, \tag{1.33}$$

де $h(q)$ - функція узагальненого показника Херста.

Із визначень та (1.31) слідує, що при $q = 2$ цей показник зводиться до звичайного значення H .

Для часових рядів, котрі відповідають монофрактальній множині, флуктуаційна функція $F_q(s)$, однакова для всіх сегментів, та узагальнений показник Херста $h(q) = H$ не залежить від параметра q . Для мультифрактальних часових рядів $h(q)$ є нелінійною функцією: при позитивних значеннях q

основних вклад в функцію $F_q(s)$ дають сегменти, котрі проявляють великі відхилення $F^2(s)$, а при від'ємних q домінують сегменти з малими дисперсіями $F^2(s)$. Таким чином, при від'ємних значеннях q узагальнений показник Херста $h(q)$ описує сегменти, котрі проявляють малі флуктуації, а при позитивних – великі.

Виходячи зі співвідношень (1.30-1.32) можна запропонувати наступний покроковий алгоритм оцінювання мультифрактальних характеристик по часовому ряду $X(t)$, $t = 1,2...,n.$

1. Для обчислення необхідно перетворити початковий ряд в кумулятивний $X^{cum}(k) = \sum_{i=1}^{k} X(t), k = 1,2,...,n.$ Якщо початковий ряд уже є кумулятивний, цей крок пропускається.

2. Задаються значення параметру q, в інтервалі яких необхідно знайти значення мультифрактальних характеристик $h(q)$, $\tau(q)$, $f(\alpha)$.

3. Ряд $X^{cum}(t)$ розбивається N неперекриваючих сегментів довжиною τ та для кожного сегменту $X^{cum}(t)$ обчислюється флуктуаційна функція

$$F^2(\tau) = \frac{1}{\tau}\sum_{t=1}^{\tau}\left(X^{cum}(t) - Y_m(t)\right)^2, \qquad (1.34)$$

де $Y_m(t)$ - локальний m-поліноміальний тренд в рамках цього сегменту.

4. Функція $F(\tau)$ усереднюється по всьому ряду $X^{cum}(t)$

$$F(\tau) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} F_i^2(\tau)}.$$

5. Обчислюється флуктуаційна функція для значення q

$$F_q(\tau) = \left\{\frac{1}{N}\sum_{i=1}^{N}\left[F^2(\tau)\right]^{\frac{q}{2}}\right\}^{\frac{1}{q}}. \qquad (1.35)$$

6. Значення τ збільшується та кроки 3-5 повторюються.

7. Будується лінійна регресія методом найменших квадратів на $\log(\tau)$, як незалежної змінної, та $\log F_q(\tau)$ як залежної змінної. Тангенс кута нахилу прямої, апроксимуючої отримані точки, визначає значення величини $h(q)$ для заданого значення q.

8. Виконавши кроки 2-6 для всіх заданих значень параметра q, отримаємо залежність $h(q)$.

9. Значення скейлінгової експоненти $\tau(q)$ визначається за формулою (1.16). Значення функції мультифрактального спектру $f(\alpha)$ знаходиться за допомогою перетворення Лежандра відповідно (1.29).

1.5 Методи машинного навчання

Основні поняття та визначення

Задано множина об'єктів X, множина допустимих відповідей Y, та існує цільова функція $y^*: X \to Y$, значення якої $y_i = y^*(x_i)$ відомі лише на кінцевій підмножині об'єктів $\{x_1,...,x_l\} \subset X$. Пари «об'єкт-відповідь» (x_i, y_i) називається прецедентами. Сукупність пар $X^l = (x_i, y_i)_{i=1}^l$ називається навчальною вибіркою.

Задача навчання прецедентами полягає в тому, щоб по вибірці X^l відновити залежність y^*, тобто побудувати вирішальну функцію $a: X \to Y$, котра наближувала б цільову функцію $y^*(x)$, причому не лише на об'єктах навчальної вибірки, але і на всій множині X.

Об'єкти та ознаки

Ознака f об'єкта x – це результат виміряння деякої характеристики об'єкта. Формально ознакою називається відображення $f: X \to D_f$, де D_f –

множина допустимих значень ознаки. Зокрема, будь-який алгоритм $a: X \to Y$ також можна розглядати як ознаку.

В залежності від природи множини D_f ознаки діляться на декілька типів:

- Якщо $D_f = \{0,1\}$, тоді f - бінарна ознака.
- Якщо D_f - кінцева множина, тоді f - номінальна ознака.
- Якщо D_f - кінцева упорядкована множина, тоді f - порядкова ознака.
- Якщо $D_f = R$, тоді f - кількісна ознака.

Якщо всі ознаки мають однаковий тип, $D_{f_1} = \cdots = D_{f_n}$, тоді вихідні дані називається однорідними, в іншому випадку – різнорідними.

Нехай є набір ознак $f_1,..., f_n$. Вектор $\left(f_1(x),..., f_n(x)\right)$ називають ознаковим описом об'єкта $x \in X$. В подальшому ми не будемо розрізняти об'єкти із X та їх ознакові описи, покладаючи $X = D_{f_1} \times ... \times D_{f_n}$. Сукупність ознакових описів всіх об'єктів вибірки X^l, записану у виді таблиці розміром $l \times n$, називають матрицею об'єктів-ознак

$$F = \left\| f_j(x_i) \right\|_{l \times n} = \begin{pmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_l) & \cdots & f_n(x_l) \end{pmatrix}$$

Типи задач

У залежності від природи множини допустимих відповідей, Y задачі навчання по прецедентам діляться на наступні типи:

- Якщо $Y = \{1,...,M\}$, то це задача класифікації на M непересічних класів. В цьому випадку вся множина об'єктів X розбивається на класи $K_y = \left\{x \in X : y^*(x) = y\right\}$, та алгоритм $a(x)$ має давати відповідь на запитання «до якого класу належить x?». У деяких прикладах класи називають образами та говорять про задачу розпізнавання образів.

- Якщо $Y = \{0,1\}^M$, то це задачі класифікації на M пересічних класів. В простішому випадку ця задача зводиться до розв'язання M незалежних задач класифікації з двома непересічними класами.
- Якщо $Y = R$, то це задача відновлення регресії.

Задачі прогнозування є окремим випадком класифікації або відновлення регресії, коли $x \in X$ – опис минулого поводження об'єкта x, $y \in Y$ - опис деяких характеристик його майбутнього поводження.

Принцип максимуму правдоподібності

При ймовірнісній постановці задачі задається модель сумісної густини розподілу об'єктів та відповідей $\varphi(x, y, \theta)$, апроксимуюча невідому густину $p(x, y)$. Потім визначається значення параметру θ, при якому вибірка даних X^l максимально правдоподібна, тобто найкращім чином узгоджується з моделлю густини.

Якщо спостереження в вибірці X^l незалежні, то сумісна густина розподілу всіх спостережень дорівнює добутку густин $p(x, y)$ в кожному спостереженні: $p(X^l) = p((x_1, y_1), ..., (x_l, y_l)) = p(x_1, y_1) \cdot ... \cdot p(x_l, y_l)$. Підставляючи замість $p(x, y)$ модель густини $\varphi(x, y, \theta)$, отримаємо функцію правдоподібності

$$L(\theta, X^l) = \prod_{i=1}^{l} \varphi(x_i, y_i, \theta)$$

Чим вище значення правдоподібності, тим краще вибірка узгоджується з моделлю. Значить необхідно шукати значення параметру θ, при якому значення $L(\theta, X^l)$ максимальне. У математичній статистиці його називають принципом максимуму правдоподібності. Його формальне обґрунтування можна знайти в [99].

Проблема перенавчання та поняття узагальнюючої здатності

Максимізацію функції правдоподібності необхідно застосовувати з відомою долею обережності. Якщо максимум функціоналу $L\left(\theta, X^l\right)$ досягається на алгоритму a, то це ще не гарантує, що a буде добре наближувати цільову залежність на довільній контрольній вибірці $X^k = \left(x_i', y_i'\right)_{i=1}^k$.

Коли якість роботи алгоритму на нових об'єктах, котрі не увійшли до навчальної вибірки, виявляється значно гірше, ніж на навчальній вибірці, говорять про ефект перенавчання [102].

Легко уявити собі метод, котрий максимізує функцію правдоподібності до одиниці, але при цьому абсолютно не здатний навчатися [99, 100]. Отримав навчальну вибірку X^l, він запам'ятовує її та будує алгоритм, котрий порівнює пред'явлений об'єкт x з навчальними об'єктами x з X^l. У випадку $x = x_i$ алгоритм видає правильну відповідь y_i, інакше видається випадкову відповідь. Функція правдоподібності приймає максимальне можливе значення, але цей алгоритм не здатний відновити залежність за межами матеріалу навчання. Звідси висновок: для успішного навчання необхідно не лише запам'ятовувати, але й узагальнювати.

Узагальнююча здатність метода μ характеризується величиною $L\left(\theta, X^l\right)$ за умови, що вибірки X^l та X^k є простими, отримані з одного і того ж самого ймовірнісного розподілу на множині X.

Випадковий ліс

Основна ідея випадкового лісу (ВЛ) [19, 102] полягає в тому, щоб оцінити велику кількість зашумлених, але приблизно неупереджених моделей, а отже, зменшити дисперсію. Дерева є ідеальними кандидатами для випадкового лісу, оскільки вони можуть охоплювати складні структури взаємодії у даних, і якщо вони виростають досить глибоко, мають відносно низький ухил. Оскільки дерева є шумливими, вони значно виграють від усереднення. Більше того, оскільки

кожне дерево, що генерується у ВЛ, розподілено однаково, математичне очікування В таких дерев збігається з математичним очікуванням будь-якого з них. Це означає, що ухил ВЛ дерев такий же, як у окремих дерев, і єдина надія на поліпшення полягає у зменшенні дисперсії.

Середнє значення незалежної однаково розподіленої випадкової величини B, кожна з дисперсією σ^2, має дисперсію $\frac{\sigma^2}{B}$. Якщо величини однаково розподілені з позитивною попарною кореляцією ρ, дисперсія має вид

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2. \qquad (1.36)$$

Зі збільшенням B другий член (1.36) зникає, але перший залишається, а значить, величина співвідношення пар ВЛ дерев обмежує переваги усереднення. Ідея у випадкових лісах полягає в поліпшенні зменшення дисперсії мішків за рахунок зменшення кореляції між деревами, не надто збільшуючи дисперсію. Це досягається в процесі вирощування дерев шляхом випадкового вибору вхідних змінних. Зазвичай значення для $n \in \sqrt{p}$ або навіть 1.

Алгоритм випадкового лісу для дерев регресії та класифікації [129]:

1. Для $b = 1,...,B$

1.1. Згенеруємо випадкову підвибірку з повтореннями Z_b розміром $N \times n$ з навчальної вибірки, де N - кількість зразків в початковій вибірці, $n = \sqrt{p}$ - кількість ознак, котрі вибираються випадково та без повторень з початкової вибірки. Таким чином, деякі зразки попадуть до неї два або більше разів, а в середньому $N\left(1-\frac{1}{N}\right)^N$, а приблизно $\frac{N}{e}$ зразків не увійдуть у вибірку. Та зразки які не попали до вибірки, називаються out-of-bag.

1.2. Виростіть дерево випадкового лісу T_b на згенерованій вибірці Z_b, рекурсивним повторенням наступних кроків для кожного кінцевого вузла дерева, поки не буде досягнуто мінімального розміру кожного кінцевого вузла дерева.

1.2.1. Виберіть змінну m серед всіх змінних p.

1.2.2. Виберіть краще значення роздільної точки серед всіх значень m.

1.2.3. Розділіть вузол на два підвузла.

2. Виведіть ансамбль дерев $\{T_b\}_1^B$.

Для регресії: $f_{rf}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$.

Для класифікації: Нехай $C_b(x)$ – є прогноз класу b-м деревом випадкового лісу. Тоді $C_{rf}^B(x) = \textit{більшість голосів}\{C_b(x)\}_1^B$.

Беггінг

Метод беггінга (bagging, bootstrap aggregation) був запропонований Л. Брейманом в 1996 році [18]. З вихідної навчальної вибірки довжиною l формуються різноманітні навчальні підвибірки тієї ж довжини l з допомогою бутстрепа – випадкового вибору з поверненням. При цьому деякі об'єкти потрапляють у вибірку по декілька разів, а деякі жодного разу. Можна сказати, що доля об'єктів, котрі потрапили в кожну підвибірку, наближається до $1-e^{-1} \approx 0,632$ при $l \to \infty$. Базові алгоритми, навчанні на підвибірках, об'єднуються в композиції за допомогою простого голосування або математичного очікування.

Ефективність беггінга пояснюється двома обставинами[130]. По-перше, завдяки різноманітності базових алгоритмів, їх похибки взаємно компенсуються під час голосування. По-друге, об'єкти-викиди можуть не потрапляти в деякі навчальні підвибірки. Тоді алгоритм, побудований за аналогією (планом) вибірки, може виявитися навіть точніше алгоритму, побудованого по повній вибірці.

Беггінг особливо ефективний на малих вибірках, коли виключення навіть невеликої долі навчальних об'єктів приводить до побудови значно різноманітних базових алгоритмів. У випадку надвеликих надлишкових вибірок доводиться будувати підвибірки меншої довжини $l_0 << l$, при цьому виникає задача вибору оптимального значення параметру l_0.

Багатошаровий персептрон

Багатошарові мережі, так само, як і одношаровий персептрон (лінійний класифікатор), можна налаштовувати градієнтними методами, незважаючи на величезну кількість вагових коефіцієнтів. У середині 80-х одночасно декількома дослідниками було запропоновано ефективний спосіб обчислення градієнта, при якому кожен градієнтний крок виконується за число операцій, лише небагато чим більше, ніж при звичайному обчисленні мережі на одному об’єкті. Це здається дивним, адже кількість операцій, необхідних для обчислення градієнта, зазвичай зростає пропорційно розмірності, тобто числу вагових коефіцієнтів. Тут цього вдається уникнути завдяки аналітичному диференціюванню суперпозиції зі збереженням необхідних проміжних величин. Метод отримав назву зворотного поширення помилок [94].

Метод зворотного поширення помилок

Розглянемо багатошарову мережу, в котрій кожен нейрон попереднього шару зв’язаний з усіма нейронами поступного шар, рисунок 1.1[56]. Така мережа називається повнозв’язною. Для більшої узагальненості покладемо $X = \mathbb{R}^n$, $Y = \mathbb{R}^M$.

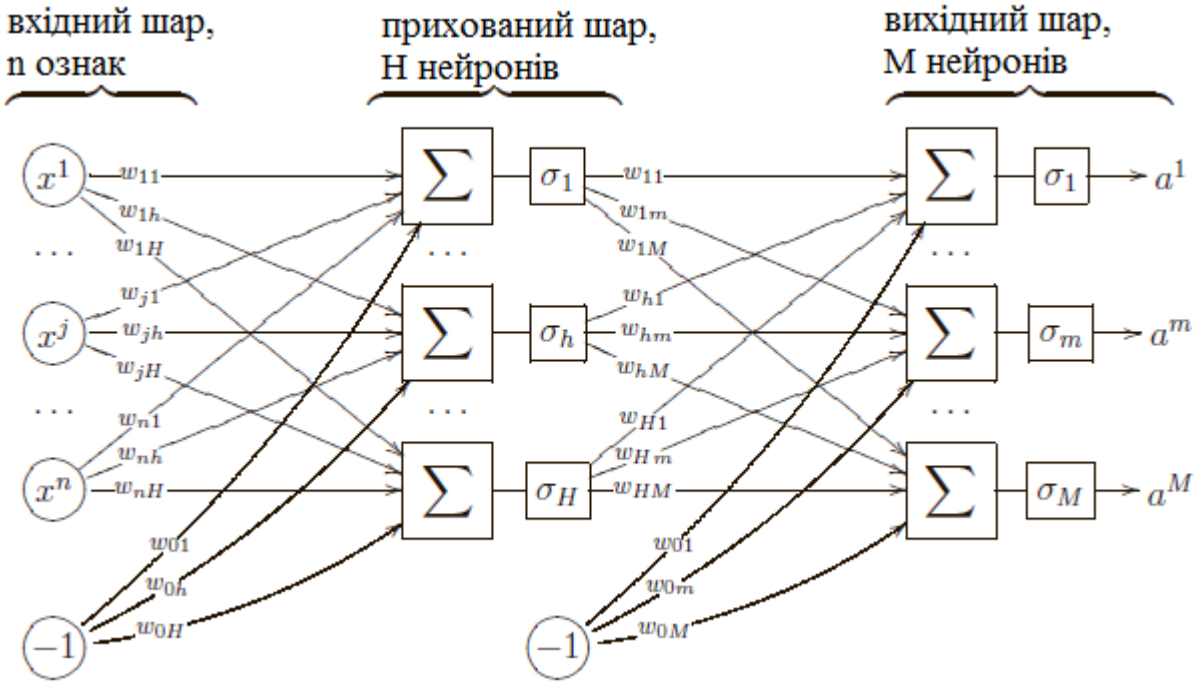

Рисунок 1.1 – Багатошарова нейронна мережа з одним прихованим шаром

Уведемо наступні позначення. Нехай вихідний шар складається зі M нейронів з функціями активації σ_m та виходами $a^m, m=1,...,M$. Перед ним знаходиться прихований шар із H нейронів з функціями активації σ^h та виходами $u^h, h=1,...,H$. Коефіцієнти синаптичних зв'язків між h-м нейроном прихованого шару та m-м нейроном вихідного шару будемо позначати через w_{hm}. Перед цим шаром може знаходитися або вхідний шар ознак, або ще один прихований шар з виходами $v^j, j=1,...,J$ та синаптичними вагами w_{jh}. У загальному випадку кількість шарів може бути будь-яким. Якщо мережа двошарова, то v^j є просто j-ю ознакою: $v^j(x) \equiv f_j(x) \equiv x^j$, та $J=n$. Позначимо через w вектор всіх синаптичних коефіцієнтів мережі.

Вихідні значення мережі на об'єкті x_i, обчислюється як суперпозиція:

$$a^m(x_i) = \sigma_m\left(\sum_{h=0}^{H} w_{hm} u^h(x_i)\right); \qquad u^h(x_i) = \sigma_h\left(\sum_{j=0}^{J} w_{jh} v^j(x_i)\right). \qquad (1.37)$$

Зафіксуємо об'єкт x_i та запишемо функціонал середньоквадратичної похибки

$$Q(w)=\frac{1}{2}\sum_{m=1}^{M}\left(a^m(x_i)-y_i^m\right)^2.$$

В подальшому нам знадобиться часткові похідні Q по вихідним нейронам. Випишемо їх спочатку для вихідного шару

$$\frac{\partial Q(w)}{\partial a^m}=a^m(x_i)-y_i^m=\varepsilon_i^m.$$

Часткова похідна Q по a^m рівна величині похибки ε_i^m на об'єкті x_i. Тепер випишемо часткові похідні по виходам прихованого шару

$$\frac{\partial Q(w)}{\partial u^h}=\sum_{m=1}^{M}\left(a^m(x_i)-y_i^m\right)\sigma'_m w_{hm}=\sum_{m=1}^{M}\varepsilon_i^m\sigma'_m w_{hm}=\varepsilon_i^h.$$

Цю величину, по аналогії з ε_i^m, будемо називати похибкою мережі на прихованому шарі та позначати через ε_i^h. Через σ'_m позначена похідна функції активації, обчислена при тому ж значенні аргументу, що і в (1.37). Якщо використовується сигмоїдна функція активації, то для ефективного обчислення похідної можна скористатися формулою $\sigma'_m=\sigma_m(1-\sigma_m)=a^m(x_i)\left(1-a^m(x_i)\right)$.

Відзначимо, що ε_i^h обчислюється по ε_i^m, якщо запустити мережу «задом наперед» (навпаки), подавши на виходи нейронів прихованого шару значення $\varepsilon_i^m\sigma'_m$, а результат ε^h отримав на вході. При цьому вхідний вектор скалярно помножується на вектори коефіцієнтів w_{hm}, котрі знаходяться праворуч від нейрона, а не ліворуч, як при прямому обчисленні.

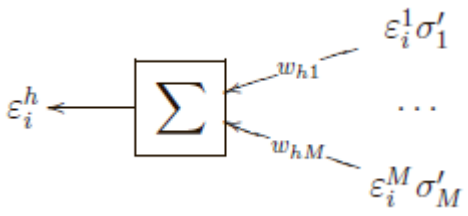

Рисунок 1.2 – Схема зворотного поширення помилки

Маючі часткові похідні по a^m та u^h, легко виписати градієнт Q по коефіцієнтам w:

$$\frac{\partial Q(w)}{\partial w_{hm}} = \frac{\partial Q(w)}{\partial a^m}\frac{\partial a^m}{\partial w_{hm}} = \varepsilon_i^m \sigma'_m u^h(x_i), \quad m = 1,...,M, \quad h = 0,...,H,$$

$$\frac{\partial Q(w)}{\partial w_{jh}} = \frac{\partial Q(w)}{\partial u^h}\frac{\partial u^h}{\partial w_{jh}} = \varepsilon_i^h \sigma'_h v^j(x_i), \quad h = 1,...,H, \quad j = 0,...,J.$$

і так далі для кожного шару. Якщо шарів більше двох, то останні часткові похідні обчислюються аналогічно – зворотнім ходом по шарам мережі справа наліво.

Переваги методу зворотного поширення похибки

- Достатньо висока ефективність. У випадку двошарової мережі прямий хід, зворотній хід та обчислення градієнта потребує порядку $O(Hn + HM)$ операцій.
- Через кожний нейрон проходить інформація тільки про зв'язаних з ним нейронах. Тому метод легко реалізується на обчислювальних пристроях з паралельною архітектурою.
- Великий ступінь узагальненості. Алгоритм легко записати для довільного числа шарів, довільної розмірності входів та виходів, довільної функції втрат та довільної функції активації, можливо різних у різних нейронах.

Недоліки методу зворотного поширення похибки

- Метод успадковує відомі недоліки градієнтного налаштування коефіцієнтів в одношаровому персептроні. Тут також виникають проблеми повільної збіжності.

1.6 Вимоги до інформаційної технології

Запропонуємо інформаційну технологію класифікації упорядкованих масивів даних із фрактальними властивостями. 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. Властивості часових рядів нескладно узагальнити на УМД, якщо замість

змінної упорядкування – часу – взяти іншу фізичну величину. Виходячи з визначення УМД, інформаційна технологія повинна класифікувати будь-які дані з єдиним обмеженням – УМД має володіти фрактальними властивостями. Наприклад інформаційна технологія має 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 та вчасної постановки діагнозу та відповідного лікування. Виявлення DDoS-атак в комп'ютерній мережі та своєчасної їх локалізації. Також ІТ має мати змогу бути застосованою для класифікації сейсмологічних графіків активності земної кори, а також для будь-яких інших УМД котрі володіють фрактальними властивостями.

Запропонована інформаційна технологія повинна збирати та зберігати вхідний потік інформації. Якщо потік вхідної інформації недостатньо активний, ІТ повинна вміти згенерувати достатню кількість інформації для проведення навчання методів машинного навчання, на основі існуючих даних та їх фрактальних характеристик. Зі збільшення об'єму інформації, інформаційна технологія повинна вміти перенавчати свої моделі класифікації з метою адаптації до зміни властивостей вхідної інформації, а також для підвищення ступеня покриття множини можливих станів вхідної інформації.

Інформаційна технологія має передавати результати класифікації деякій зовнішній системі котра використовує результати роботи запропонованої інформаційної технології. Інформаційна технологія повинна отримувати від зовнішньої системи результати використання класифікації, та проводити аналіз ефективності роботи інформаційної технології.

Навчена модель класифікації зберігається в базі даних, та використовується для класифікації вхідного потоку інформації. До тих пір поки модель ще не навчена ІТ лише збирає та зберігає вхідну інформацію.

Таким чином можна формалізувати вимоги до інформаційної технології:

- 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.

1.7 Постановка задачі дослідження

Метою дисертаційної роботи є створення інформаційної технології класифікації упорядкованих масивів даних із фрактальними властивостями методами машинного навчання для підвищення точності класифікації.

Згідно з зазначеною метою треба розв’язати такі задачі:

– провести аналіз існуючих методів класифікації упорядкованих масивів даних методами машинного навчання;

– розробити методи генерації упорядкованих даних з фрактальними властивостями різних типів для побудови навчальних вибірок для виконання навчання моделі класифікації з вчителем;

– виконати експериментальні дослідження розроблених моделей даних та методів класифікації для визначення їх ефективності;

– запропонувати інформаційну технологію класифікації фрактальних упорядкованих масивів даних

ВИСНОВКИ ПО ПЕРШОМУ РОЗДІЛУ

1. Багато складних систем мають фрактальну структуру, а їх динаміка представлена упорядкованими масивами даних, які мають фрактальні властивості. Аналіз фрактальних властивостей широко використовується в різних областях знань. У багатьох випадках виникають проблеми розпізнавання і класифікації фрактальних даних: в телекомунікаціях, для виявлення кібератак; в геології – для розподілення та класифікації сейсмічних хвиль; у біології та медицині – для діагностики захворювань і фізіологічного стану; в економіці – для прогнозування кризових ситуацій за фінансовими рядами.

2. Класичний підхід до класифікації фрактальних даних базується на оцінюванні фрактальних характеристик та їхніх довірчих інтервалів, на основі яких проводиться прийняття рішення про належність даних до певного класу. При оцінюванні показника Херста найбільш популярними є методи нормованого розмаху і детрендованного флуктуаційного аналізу. При оцінюванні мультифрактального характеристик найбільш затребуваним є метод мультифрактального детрендованного флуктуаційного аналізу

3. В останні роки зростає застосування методів машинного навчання для аналізу і класифікації фрактальних масивів даних. Методи машинного навчання дозволяють класифікувати дані, базуючись лише на їх статистичних характеристиках, або не вимагають високої точності оцінки статистичних та фрактальних характеристик. Використання методів машинного навчання дозволяє компенсувати похибки оцінювання фрактальних статистичних характеристик при побудові класифікатора.

4. Проведений аналіз існуючих методів класифікації упорядкованих масивів даних методами машинного навчання показав: незважаючи на те, що застосування машинного навчання в задачах аналізу фрактальних даних постійно зростає, на сьогоднішній, день не розроблено інформаційної технології, яка призначена для задач класифікації фрактальних масивів даних, зокрема часових рядів.

Результати першого розділу опубліковані в [59, 62, 106, 117, 122, 123].

РОЗДІЛ 2
ВИКОРИСТАННЯ МОДЕЛЮВАННЯ ФРАКТАЛЬНИХ УПОРЯДКОВАНИХ МАСИВІВ ДАНИХ ЯК МЕТОД ДОПОВНЕННЯ ДАНИХ ДЛЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ В УМОВАХ НЕДОСТАТНЬОЇ ІНФОРМАЦІЇ

Для задоволення вимоги про недостатність реальних даних для навчання моделі класифікації, необхідно дослідити методи генерації фрактальних процесів на основі фрактальних та мультифрактальних властивостей упорядкованих масивів даних та визначити найбільш оптимальні методи в залежності від типу упорядкованого масиву даних (УМД). В роботах [11, 12, 30, 31, 41, 72] представлені методи генерації УМД з фрактальними та мультифрактальними властивостями.

У 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. В 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. В таких ситуаціях зазвичай використовують методи математичного моделювання, котрі дозволяють згенерувати необхідну кількість навчальних прикладів з 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:

$$Y(p) = f(process,\ parameters,\ N),\ parameters = g\ (estimates\ (X\left(q\right), N_X)),$$

де: $Y(p)$, $p = 1,...,N$, – модельна реалізація необхідної довжини N; *process* - базовий стохастичний фрактальний процес; *parameters* - параметри цього

процесу; $X(q)$ – досліджуваний УМД довжини N_X, $q=1,...,N_X$; f,g - деякі перетворення, *estimates* - статистичні та фрактальні характеристики, отримані по УМД $X(q)$.

2.1 Біноміальний стохастичний мультиплікативний каскад

Мультиплікативні каскади вперше були введені А.М. Колмогоровим для опису нерегулярного розділення швидкості дисипації енергії турбулентного потоку. Далі вивчення та розвиток каскадних моделей було представлено працями Б. Мандельброта. У сучасний час мультиплікативні каскади використовуються для моделювання телекомунікаційних трафіків, процесу випаду осадків, динаміки фінансових ринків.

Побудова біноміальних стохастичних мультиплікативних каскадів. Розглянемо процедуру побудови консервативного стохастичного каскада, яка бере основу з двохосновних частин: перша передбачає нескінченну геометричну деталізацію шляхом ітераціонального розбиття інтервалів, а друга забезпечує випадковість вагових коефіцієнтів [109].

Для кожного фіксованого значення t існує єдина послідовність $k_1,k_2,...$ така, як двійковий інтервал $I_{k_n}^{(n)}=\left[k_n 2^{-n},\left(k_n+1\right)2^{-n}\right]$ містить t для всіх цілих n, для яких $k_n=k_n(t)=\left\lfloor t\cdot 2^n \right\rfloor$. Очевидно, що будь-яке t визначає таким чином єдину вкладену послідовність $\{k_n\}$, таку, що $I_{k_{n+1}}^{(n+1)}\subset I_{k_n}^{(n)}$.

Основна ідея біноміальної міри – перерозподіл маси інтервалу $I_{k_n}^{(n)}$ на два двійкових підінтервала $I_{2k_n}^{(n+1)}$ та $I_{2k_n+1}^{(n+1)}$ пропорційно заданим числам (множникам) $M_{2k_n}^{(n+1)}$ та $M_{2k_n+1}^{(n+1)}$. Для узгодженості необхідно щоб $M_{2k_n}^{(n+1)}+M_{2k_n+1}^{(n+1)}=1$.

При визначених множниках $M_{k_n}^{(n)}$ біноміальна міра μ_b інтервалу $I_{k_n}^{(n)}$ визначається наступним чином

$$\mu_b\left(I_{k_n}^{(n)}\right)=M_{k_n}^{(n)}\cdot M_{k_{n-1}}^{(n-1)}\cdot ...\cdot M_{k_1}^{(1)}\cdot M_{k_0}^{(0)}, \qquad (2.1)$$

Біноміальний каскад K_b – це процес, який стартує з 0: $K_b(0)=0$ та виконується

$$K_b\left((k_n+1)2^{-n}\right)-K_b\left(k_n 2^{-n}\right)=\mu_b\left(I_{k_n}^{(n)}\right). \tag{2.2}$$

Кожна реалізація біноміального каскада є функцією розділення, зокрема, кожна реалізація – неперервна справа та зростаюча. Майже всі траєкторії не мають похідної. На рис. 2.1 показані 4 ітерації побудови детермінованого біноміального каскада (згори). В нижній частині рис. 2.1 зліва показано прирощення каскадного процесу, отримані при числі ітерацій $k=10$, справа – реалізація каскадного процесу K_b.

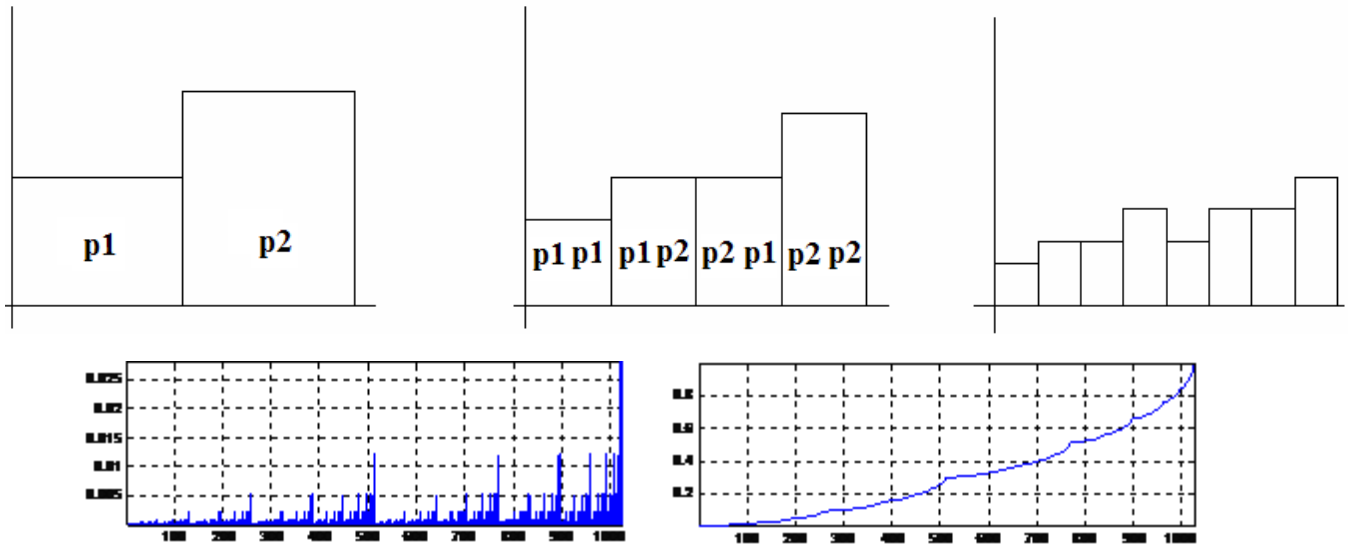

Рисунок 2.1 – Ітерації побудови каскада (зверху), реалізація прирощення (внизу зліва), реалізація каскада

Для генерації випадкового каскадного процесу множники $M_{k_n}^{(n)}$ повинні бути випадковими величинами, розподіл ймовірностей яких повинен задовольняти наступні вимоги:

1) Будь-яка підходяща позитивна випадкова величина може бути вибрана в якості $M_0^{(0)}$.

2) Величини $M_k^{(n)}$ повинні бути позитивними майже напевно та $M_{2k_n}^{(n+1)} + M_{2k_n+1}^{(n+1)} = 1 \ \forall n,k.$

3) Всі множники вкладеної послідовності повинні бути взаємно незалежними випадковими величинами. З цього слідує, враховуючі (2.2)

$$M\left[M_{k_n}^{(n)} \cdot M_{k_{n-1}}^{(n-1)} \cdot \ldots \cdot M_0^{(0)}\right] = M\left[M_{k_n}^{(n)}\right] \cdot M\left[M_{k_{n-1}}^{(n-1)}\right] \cdot \ldots \cdot M\left[M_0^{(0)}\right],$$

$$\forall n,k \ \mathrm{Low}\left\{M_k^{(n)}\right\} = \begin{cases} \mathrm{Low}\left\{M_0\right\}, \text{якщо } k - \text{парне}, \\ \mathrm{Low}\left\{M_1\right\}, \text{якщо } k - \text{непарне}. \end{cases}$$

Таким чином, при побудові консервативних стохастичних каскадів ваговими коефіцієнтами є незалежні значення деякої заданої позитивної випадкової величини W. Випадкова величина вибирається таким чином щоб сума вагових коефіцієнтів на кожній ітерації дорівнювала одиниці(вимога 2). Якщо вибрати випадкову величину, визначену на інтервалі $[0,1]$, то сума коефіцієнтів на кожній ітерації буде дорівнювати одиниці.

У цьому випадку першим двом інтервалам будуть приписані вагові коефіцієнти w_1 та $1-w_1$ відповідно. Під час другого кроку додаються два нових незалежних значення w_2 та w_3. Виходить чотири інтервали з ваговими коефіцієнтами $w_1 w_2$, $w_1(1-w_2)$, $(1-w_1)w_3$ та $(1-w_1)(1-w_3)$. При $n \to \infty$ досягаємо граничної міри що являється неоднорідною фрактальною множиною.

В роботі [111] запропоновано в якості випадкової величини, що породжує вагові коефіцієнти, використати випадкову величину, що має бета-розподіл. Бета-розподіл з параметрами $a > 0$, $b > 0$, називається розподілом з густиною ймовірностей:

$$p_x(\alpha,\beta) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & x \in (0,1), \\ 0, & x \notin (0,1), \end{cases} \tag{2.3}$$

де $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ - бета-функція.

Форма графіків густини розподілу ймовірності випадкової величини з бета-розподілом залежить від вибору параметрів a та b. На рис. 2.2 зліва представлені графіки густини розподілу для симетричного бета-розподілу ($a=b$) при значеннях $a=\{0,5;\ 1;\ 1,5;\ 3\}$. При значеннях параметрів $a=b=1$ ми отримаємо випадкову величину, котра має рівномірний розподіл на інтервалі $[0;1]$. На рис. 2.2 справа приведені графіки несиметричних розподілів.

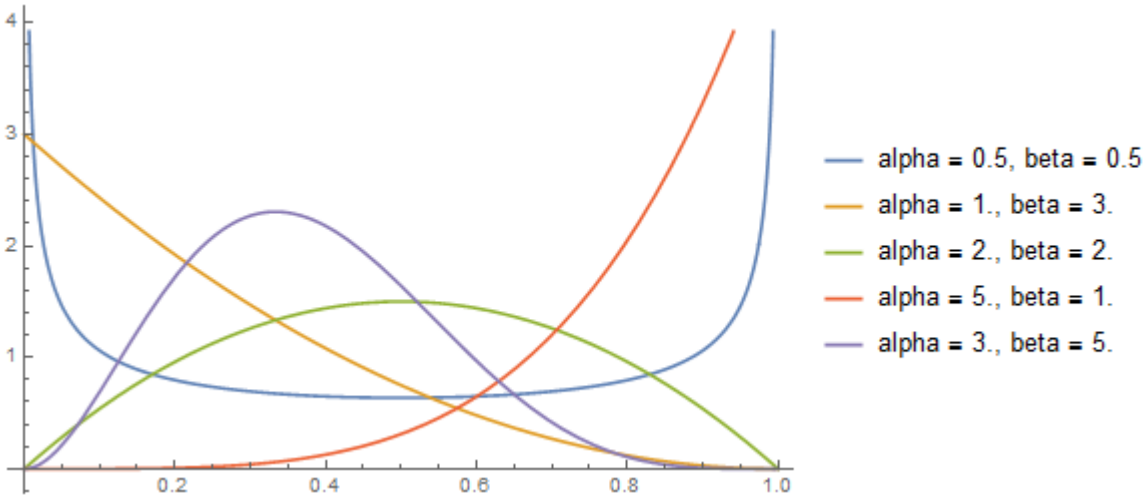

Рисунок 2.2 – Густини розподілу для симетричного та несиметричного бета-розподілу

Математичне очікування та дисперсія випадкової величини з бета-розподілом визначається наступним чином:

$$M[X]=\frac{\alpha}{\alpha+\beta},\ D[X]=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{2.4}$$

Моменти $q-$го порядку визначаються по формулі

$$M\left[X^q\right]=\frac{B(\alpha+q,\beta)}{B(\alpha,\beta)}. \tag{2.5}$$

Для біноміальних мультиплікативних стохастичних каскадів при виконанні умов 1) - 3) скейлінгова експонента $\tau(q)$ зв'язана з множниками $M_k^{(n)}$ наступним співвідношенням

$$\tau(q)=-\log_2 M\left[(M_0)^q+(M_1)^q\right]-1. \tag{2.6}$$

У рамках вищеописаного алгоритму побудови на кожному кроці величина $M_{2k_n}^{(n+1)}$ має бета-розподіл, а величина $M_{2k_n+1}^{(n+1)} = 1 - M_{2k_n}^{(n+1)}$. У цьому випадку вираз (2.6), враховуючи властивість 2) зводиться до виду

$$\tau(q) = -\log_2 M\left[\left(M_0\right)^q + \left(1 - M_0\right)^q\right] - 1. \tag{2.7}$$

Для бета-розподілу, враховуючи формулу вищих моментів (2.5) та властивості бета-функції, можна аналітично визначити скейлінгову експоненту $\tau(q)$

$$\tau(q) = -\log_2 \frac{B(\alpha + q, \beta)}{B(\alpha, \beta)} - 1, \tag{2.8}$$

для значень $q > -\alpha$.

На рис. 2.3 представлені 9 ітерацій побудови біноміального каскада з рівномірним розподілом вагових коефіцієнтів.

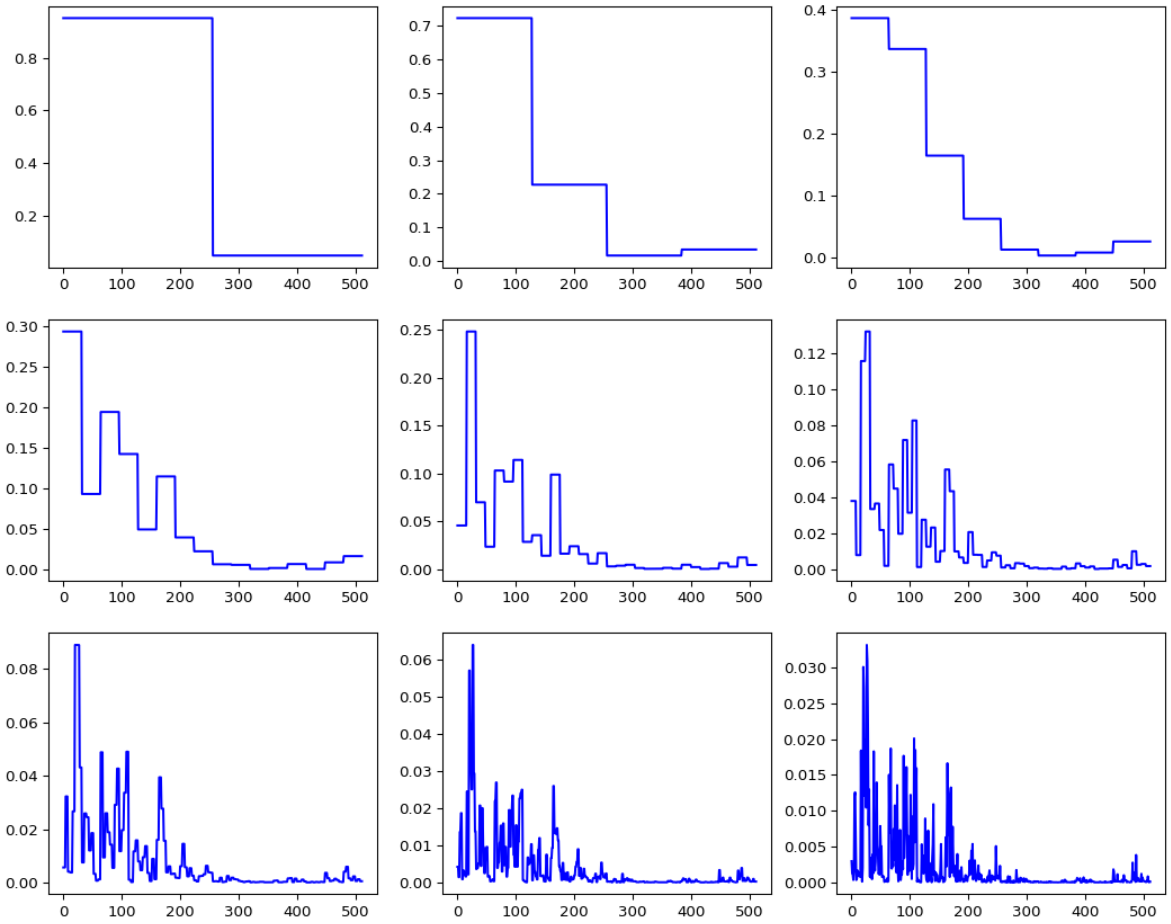

Рисунок 2.3 – Ітерації побудови каскада з рівномірним розподілом вагових коефіцієнтів, та показником Херста $H = 0.85$

2.2 Моделювання біноміальних стохастичних мультиплікативних каскадів, на основі оцінок мультифрактальних характеристик часових рядів

У загальному випадку модельна реалізація , побудована на основі часового ряду, котрий володіє фрактальними властивостями, може бути представлена наступним чином[98, 103, 105, 126, 128, 131]:

$$Y(t) = f(process,\ parameters,\ N),\ parameters = g\ (estimates\ (X, N_X))$$

де $Y(t)$, $t = 1,...,N$ - модельна реалізація необхідної довжини N; *process* - базовий фрактальний процес з дискретним часом; *parameters* - параметри цього процесу; $X(t)$ - досліджуваний часовий ряд довжини N_x, $t = 1,...,N_x$; *estimates* - статистичні характеристики, отримані по часовому ряду $X(t)$; f, g - деякі перетворення.

У розглянутому випадку базовим фрактальним процесом *process* є біноміальний консервативний мультиплікативний каскад з ваговими коефіцієнтами, котрі є значеннями випадкової величини з бета-розподілом; параметри *parameters*, потребують визначення є параметри α та β бета-розподілу. Необхідно визначити α та β так, щоб скейлінгова експонента $\tau(q)$ каскадного мультифрактального процесу співпадала з оцінкою скейнлінгової експоненти по часовому ряду $X(t)$.

Був запропонований механізм визначення параметрів α та β несиметричного бета-розподілу для побудови мультиплікативного каскаду з заданими мультифрактальними характеристиками з діапазоном показника Херста $H \in (0.5;1.0)$. Параметри симетричного бета-розподілу визначаються лише показником Херста та можуть бути визначені розв'язанням рівнянням (2.8) при $q = 1, \alpha = \beta$ з використанням більш простої залежності для другого початкового моменту, отриманого за допомогою функції твірної моментів $M\left[X^2\right] = \dfrac{\alpha(1+\alpha)}{(\alpha+\beta)(1+\alpha+\beta)}$. Несиметричний бета-розподіл дозволяє генерувати мультиплікативний каскад з більш схожою динамікою змінення функції $\tau(q)$, що

в свою чергу дозволяє охопити більш великий спектр мультифраактальних властивостей.

Для побудови каскаду з заданими мультифрактальними властивостями необхідно розв'язати рівняння (2.8) відносно параметрів α та β. Дане рівняння не має аналітичного розв'язання відносно невідомих параметрів, оскільки функція $B(\alpha,\beta)$ не має аналітичної зворотної функції, тому ми можемо приблизити точне рішення методом найменших квадратів [132], сформулювавши наступну задачу нелінійної умовної оптимізації [107]

$$\alpha,\beta = \arg\min_{\alpha,\beta}\left(\tau(q)+\log_2\frac{B(\alpha+q,\beta)}{B(\alpha,\beta)}+1\right)^2, \qquad (2.9)$$
$$\alpha>0,\ \beta>0,\ q>-\alpha.$$

Для проведення чисельних досліджень з реальними даними, були відібрані реалізації мережних трафіків різних протоколів, котрі володіють фрактальними властивостями. На рисунку 2.4 приведений графік вибіркової реалізації трафіка довжиною 8182 точки.

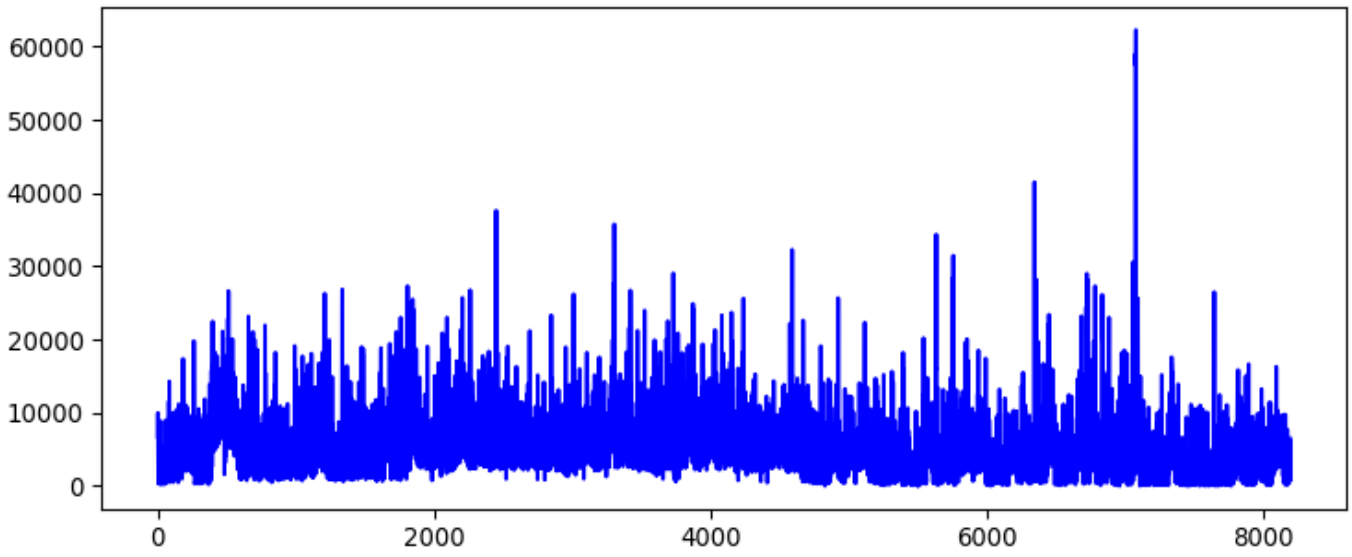

Рисунок 2.4 – Реалізація мережного трафіку

Обчислена за допомогою метода МФДФА скейлінгова експонента $\hat{\tau}(q)$ представлена на рисунку 2.5. Нелінійність скейлінгової експоненти вказує на наявність мультифрактальних властивостей трафіку.

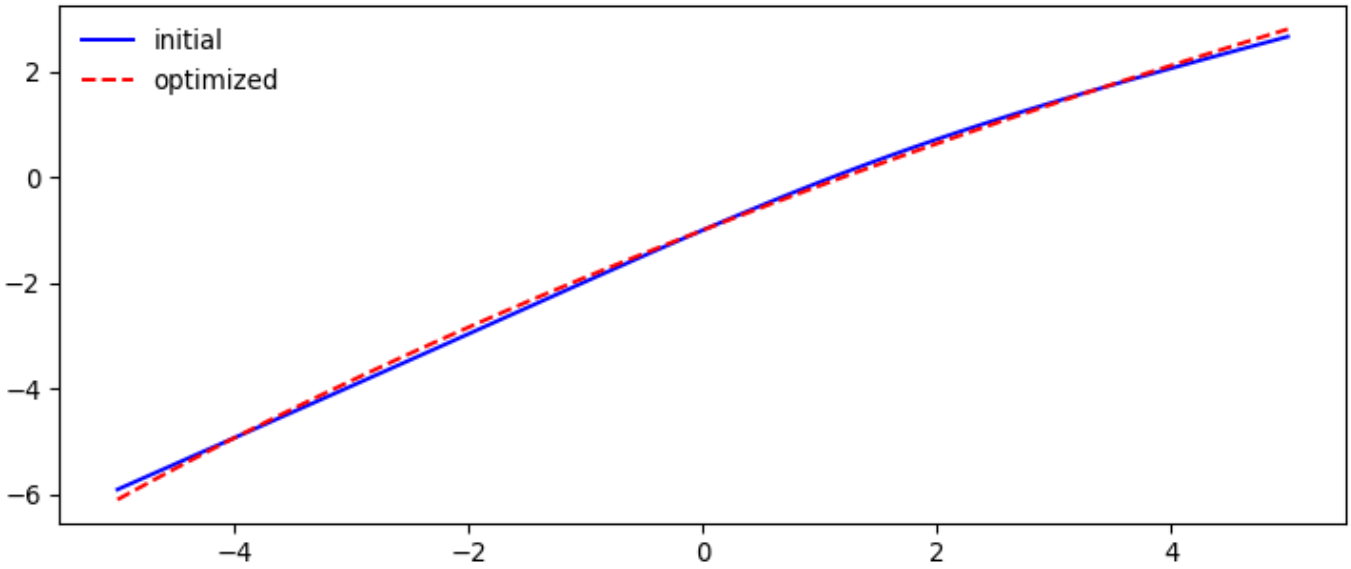

Рисунок 2.5 – Скейлінгова експонента $\tau(q)$: «initial» - оцінка по ряду трагіку, «optimized» - розв'язок оптимізаційної задачі (2.9)

При розв'язанні оптимізаційної задачі (2.9) були отримані параметри бета-розподілу, котрі відповідають оціненій функції $\hat{\tau}(q)$ $\alpha = 13,64846$ и $\beta = 10,80131$. На рисунку 6 представлений графік густини отриманого бета-розподілу.

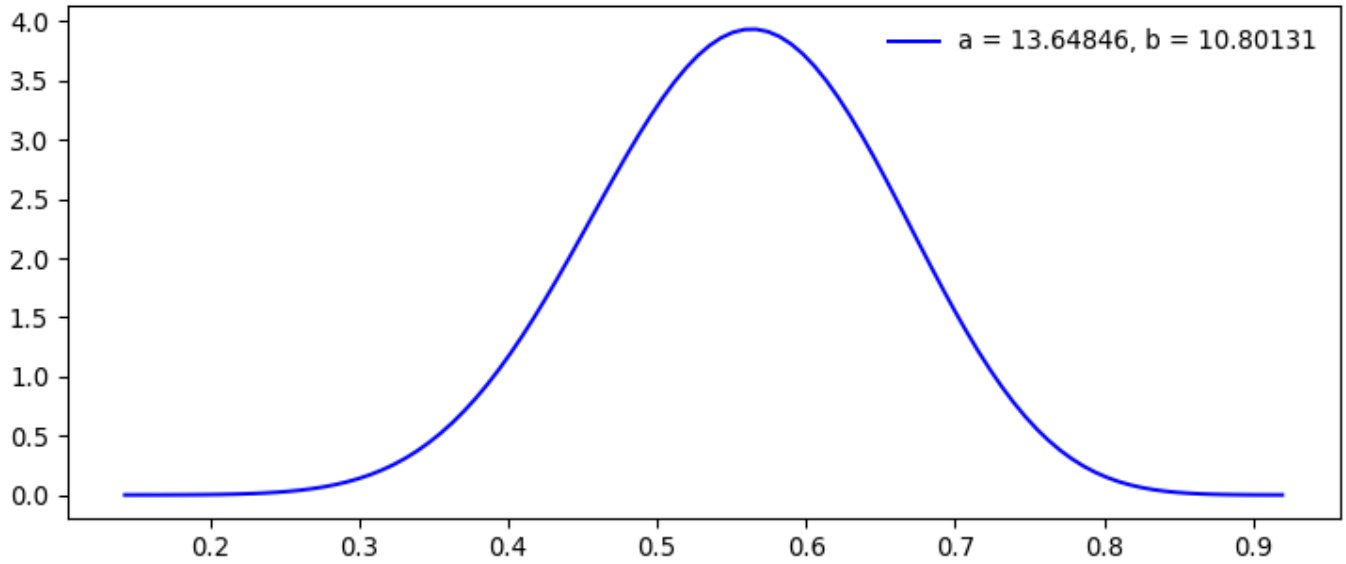

Рисунок 2.6 – Трафік густини бета-розподілу мультиплікативного каскаду

Теоретична скейлінгова експонента $\tau(q)$ для біноміального мультиплікативного каскаду представлена на рисунку 2.5. Одна з можливих модельних реалізацій каскаду з такими мультифрактальними властивостями представлена на рисунку 2.7.

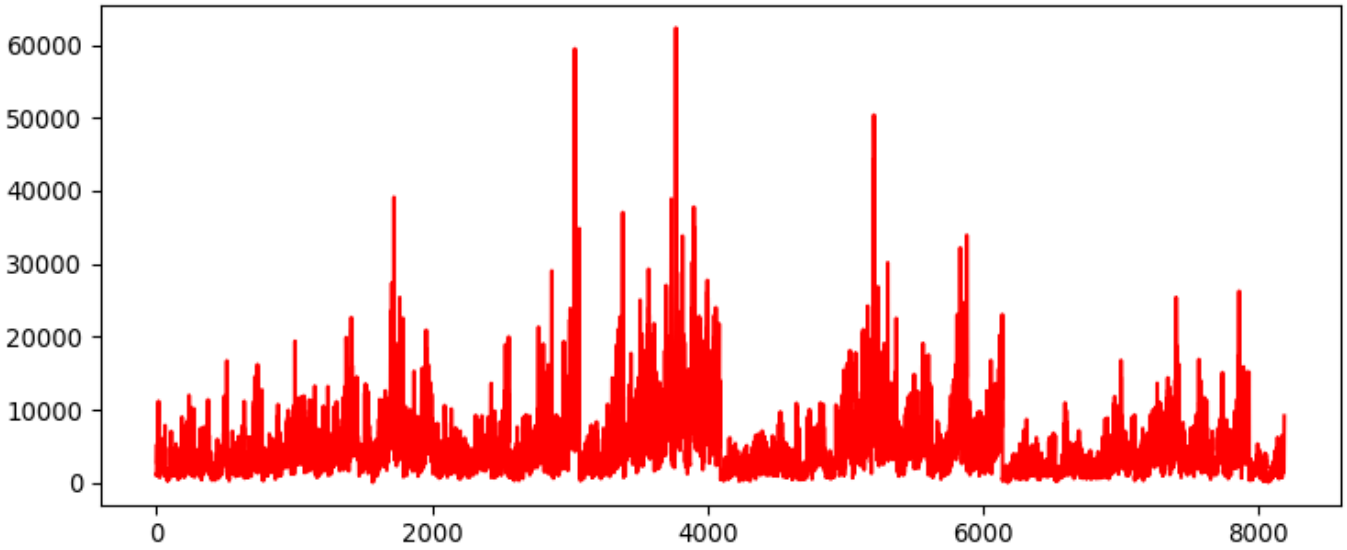

Рисунок 2.7 – Реалізація мультифрактального каскаду з параметрами бета-розподілу $\alpha = 13,64846$ та $\beta = 10,80131$.

2.3Експоненціальне перетворення фрактального гаусівського шуму

Розглянемо добре відомий фрактальний процес броунівського руху[6, 8, 109]. Гаусівський процес $X(t)$ називається фрактальним броунівським рухом за параметром $H, 0 < H < 1,$ якщо прирощення випадкового процесу $\Delta X(\tau) = X(t+\tau) - X(t)$ має гаусівське розподілення: $P(\Delta X < x) = \frac{1}{\sqrt{2\pi}\sigma_0 \tau^H} \cdot \int_{-\infty}^{x} \mathrm{Exp}\left[-\frac{z^2}{2{\sigma_0}^2 \tau^{2H}} \right] dz$, де σ_0 - коефіцієнт дифузії.

Фрактальний броунівський рух з параметром $H = 0.5$ співпадає з класичним броунівським рухом [83]. Прирощення ФБР називаються фрактальним гаусівсьвим шумом(ФГШ). Дисперсія ФГШ підкорюється відношенню $D[X(t+\tau) - X(t)] = {\sigma_0}^2 \tau^{2H}$.

Прирощення фрактального броунівського руху ΔX володіє властивістю самоподібності у вузькому сенсі, тобто випадкові процеси $X(t+\tau) - X(t)$ та $a^{-H}\left(X(t+a\tau) - X(t)\right)$ мають однакові кінцевомірні розподіли для будь-якого $a > 0$.

Теорема. Прирощення фрактального броунівського руху ΔX володіє властивістю самоподібності у вузькому сенсі, тобто $X(t+\tau)-X(t) \triangleq a^{-H}\left(X(t+a\tau)-X(t)\right)$ для будь-якого $a>0$.

Теорему та її доведення розглянуто в [121].

Існує декілька методів побудови ФБР для випадку дискретного часу. Одним з найбільш використовуваним на практиці є метод послідовного випадкового складання Фосса, що включає в себе наступний покроковий алгоритм [128]. Вихідні значення координат $X(t_i)$ в момент часу $t_i = 0,\ 1/2,\ 1$ дорівнюють нулю. На першому кроці до значення координат $X(t_1)$, $X(t_2)$, $X(t_3)$ додаються випадкові числа, вибрані із нормального розподілу з нульовим середнім і початковою дисперсією σ_1^2. Середнє значення часу на кожному інтервалі розглядається як додаткові вузли на осі часу; значення координат в них оцінюється інтерполяцією. Під час наступного кроку до всіх координат $X(t_i)$, $(t_i = 0,\ 1/4,\ 1/2,\ 3/4,\ 1)$ прибавляються випадкові числа з нульовим середнім значенням та зменшеної дисперсії $\sigma_2^2 = \frac{\sigma_1^2}{2^{2H}}$. На n-му кроку алгоритму ми отримуємо значення реалізації ФБР для $1+2^n$ значень часу t_i. Дисперсія складових n-го покоління дорівнює

$$\sigma_n{}^2 = \frac{\sigma_{n-1}{}^2}{2^{2H}} = \frac{\sigma_0{}^2}{2^{2Hn}}. \qquad (2.10)$$

Процес, запропонований Фоссом, приводить до узагальненого броунівського руху при любому масштабі. На рисунку 2.8 представлені реалізації ФБР, отримані методом Фосса.

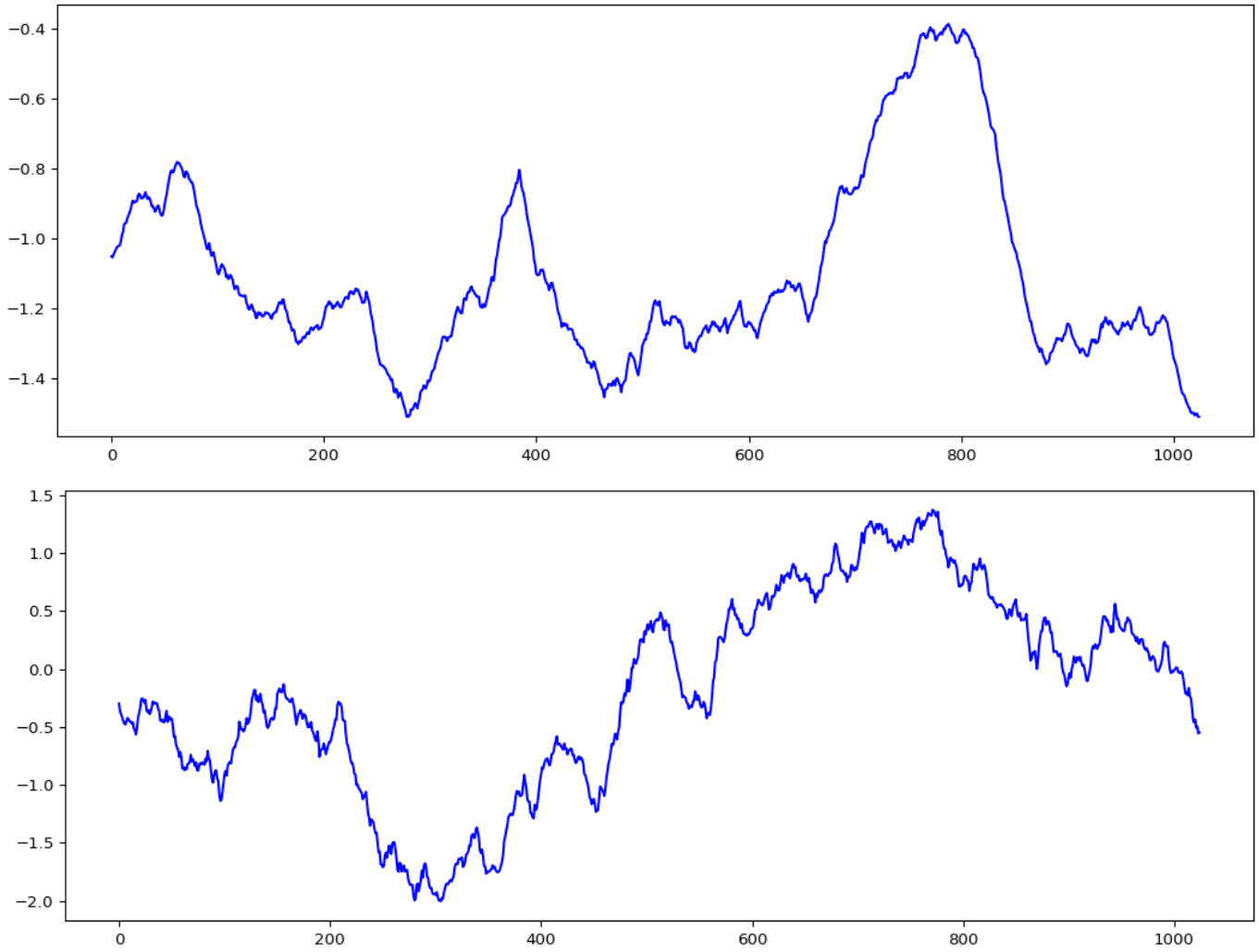

Рисунок 2.8 – Представлені реалізації ФБР, отримані методом Фосса. Зверху $H = 0{,}9$, знизу $H = 0{,}6$

Теоретично ФГШ є моделлю самоподібного процесу з деяким показником Херста та з відповідною довгостроковою залежністю. Однак ця модель володіє суттєвим недоліком: відсутністю важких хвостів, характерних для багатьох природних процесів. На практиці найчастіше використовуються розподілення з важкими хвостами розподілення Парето, Коши, Леви та логарифмічно нормальне розподілення.

Величина η має логарифмічно нормальне розподілення $\mathrm{Log}N(a;\sigma)$, якщо $\eta = \mathrm{Exp}[\xi]$, де випадкова величина ξ розподілено нормально $N(a;\sigma)$. Густина випадкової величини η

$$p_{\eta}(y)=\begin{cases}\dfrac{1}{y\sqrt{2a\sigma}}e^{-\frac{(\ln y-a)^2}{2\sigma^2}}, y>0,\\ 0, y\le 0.\end{cases} \tag{2.11}$$

Для випадкової величини з логнормальним розподіленням математичне очікування та дисперсія відповідно рівні:

$$\begin{aligned} M[\eta]&=e^{a}\cdot e^{\frac{\sigma^2}{2}},\\ D[\eta]&=e^{a^2}\cdot e^{\sigma^2}\cdot\left(e^{\sigma^2}-1\right). \end{aligned} \tag{2.12}$$

На рисунку 2.9 приведені графіки густини логнормального розподілення з різними параметрами a та σ. Важкість хвостів $P[\xi>x]$ залежить від відношення $F[\eta]=\dfrac{D[\eta]}{M[\eta]}$, яке називається індексом дисперсії або фактором Фано. Таким чином, величина $F[\eta]$ являється кількісною характеристикою важкості хвоста логнормального розподілення.

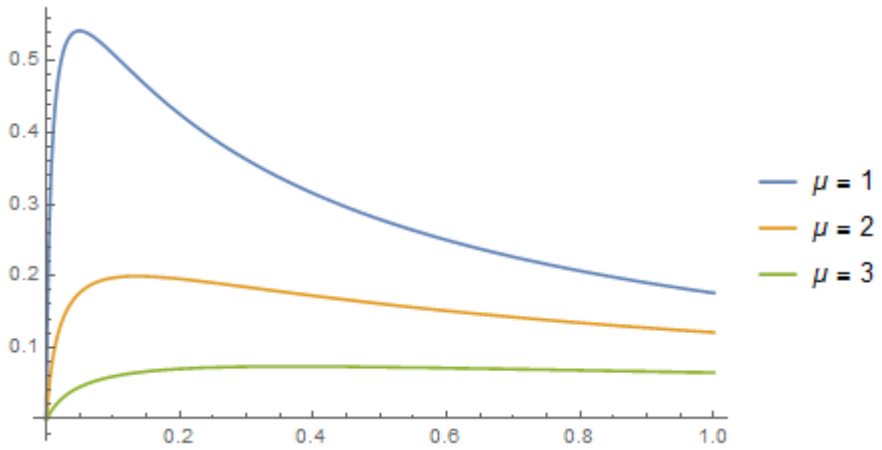

Рисунок 2.9 – Густини логарифмічно нормальних розподілень

В роботі [109] запропонований підхід, який заснований на функціональному перетворенні ФГШ. Запропоноване перетворення зберігає довгострокову залежність випадкового процесу та переводе його в самоподібний процес з важкими хвостами. В якості випадкового процесу, який породжує модельні

реалізації з заданими характеристиками самоподібності та важкістю хвостів, запропоновано використовувати наступне функціональне перетворення ФГШ

$$Y(\tau) = b \cdot \mathrm{Exp}\left[k \cdot X(\tau)\right], \tag{2.13}$$

де $X(\tau)$ - ряд прирощень ФГР з заданим параметром H на інтервалі часу τ, b, k - параметри, котрі регулюють важкість хвостів розподілення.

Пророщення ФБШ $X(\tau)$ мають нормальне розподілення $X \cong N\left(0, \sigma^2(\tau)\right)$, з дисперсією $\sigma^2(\tau) = \sigma_0^2 \tau^{2H}$, де σ_0 - коефіцієнт дифузії. У цьому випадку $Y(\tau)$ має логарифмічне нормальне розподілення $Y \cong \mathrm{Log}\, N\left(0, \sigma^2\right)$.

Параметри b, k визначають за формулами:

$$\begin{aligned} b &= \frac{M[Y(\tau)]}{\sqrt{\dfrac{F[Y(\tau)] + M[Y(\tau)]}{M[Y(\tau)]}}}, \\ k &= \frac{\sqrt{\ln\left(\dfrac{F[Y(\tau)] + M[Y(\tau)]}{M[Y(\tau)]}\right)}}{\sigma(\tau)}, \end{aligned} \tag{2.14}$$

де $\sigma^2(\tau)$ - дисперсія ряду ФГШ $X(\tau)$.

Під час генерації ФБР методом Фосса величина $\sigma^2(\tau) = \sigma_0^2 \tau^{2H}$ представляє собою дисперсію різності випадкових величин n-го покоління та залежить від довжини модельованого часового ряду та показника Херста. Із співвідношення (2.10), враховуючи, що початкова дисперсія $\sigma_0^2 = 1$ та кількість точок першого покоління дорівнює 3, отримуємо:

$$\sigma^2(\tau) = \left(\frac{1}{2}\right)^{2H(n-3)}, \tau = 2^n. \tag{2.15}$$

2.4Мультифрактальний броунівський рух. Побудова реалізацій фрактального броунівського руху в мультифрактальному часі на основі каскадної моделі

Безпосереднє моделювання каскадних реалізацій з заданими мультифрактальними параметрами приводить до процесів з корельованим додатнім прирощенням. Такими властивостями володіють реалізації інформаційних трафіків чи динаміка волатильності фінансових ринків. Однак більшість фінансових рядів володіють прирощеннями різного знаку.

В 1996р. Б. Мандельброт запропонував мультифрактальну модель, котра описує динаміку фінансових рядів MMAR (Multifractal Model of Asset Returns) [25]. Модель враховує емпіричні закономірності фінансових часових рядів, такі, як важкі хвости та довгострокову пам'ять. Вона заснована на моделюванні фрактального броунівського руху в мультифрактальному часі за допомогою операції субординації [29].

У загальному випадку операція субординації (випадкова заміна часу) подана в ідеї

$$Z(t) = Y(T(t)),\ t \geq 0, \tag{2.16}$$

де $T(t)$ - деякий негативний не спадаючій випадковий процес, котрий називається субординатором, $Y(t)$ - деякий випадковий процес незалежний від процесу $T(t)$.

У роботі [25] доведено наступне твердження. Нехай $X(t)$ є процесом субординації

$$X(t) = B_H(\theta(t)), \tag{2.17}$$

де $B_H(t)$ - фрактальний броунівський рух з параметром Херста H, $\theta(t)$ - субординатур, котрий є функцією розподілу мультифрактальної кумулятивної міри, визначеної на інтервалі $[0,T]$ та процеси $B_H(t)$ та $\theta(t)$ є незалежними.

Тоді $X(t)$ є мультифрактальних процесом з наступними характеристиками

$$\tau_X(q) = \tau_\theta(Hq),\ f_X(\alpha) = f_\theta\left(\frac{\alpha}{H}\right), \tag{2.18}$$

де $\tau_\theta(q)$ та $f_\theta(\alpha)$ - мультифрактальні характеристики процесу $\theta(t)$.

У даному випадку алгоритм моделювання реалізації мультифрактального процесу з заданими характеристиками може бути зведений до наступної процедури:

1. Для процесу $X(t)$, котрий володіє властивостями мультифрактальності, по реалізації методами мультифрактального аналізу, які розглянуті в розділі 1, оцінюємо функцію мультифрактального спектру $\hat{f}_X(\alpha)$, $\hat{\tau}_X(q)$ та параметр самоподібності $\hat{H}_X$.
2. По формулі (2.18) визначаємо мультифрактальні характеристики $\tau_\theta(q)$ та $f_\theta(\alpha)$ стохастичного каскадного процесу $\theta(t)$.
3. По табличним даним вибираємо параметри a та b, котрі визначають закон бета-розподілу $p_x(a,b)$, що генерує вагові коефіцієнти мультифрактального каскаду з характеристиками $\tau_\theta(q)$ та $f_\theta(\alpha)$.
4. За допомогою ітераційної процедури, описаної в цьому розділі, будуємо реалізацію біноміального мультиплікативного каскаду $\theta(t)$, $t = 1,2,...,2^n$, на основі $p_x(a,b)$ необхідної довжини 2^n, де n - кількість ітерацій.
5. За формулою (2.17) генеруємо реалізацію $X(t)$.

На рисунку 2.10 приведено порівняння реальних даних годинних котирувань валютної пари EURUSD з 01.08.2019 по 01.08.2020 та модельних даних мультифрактального броунівського руху довжиною 8192 значення побудованих по фрактальним характеристикам котирувань.

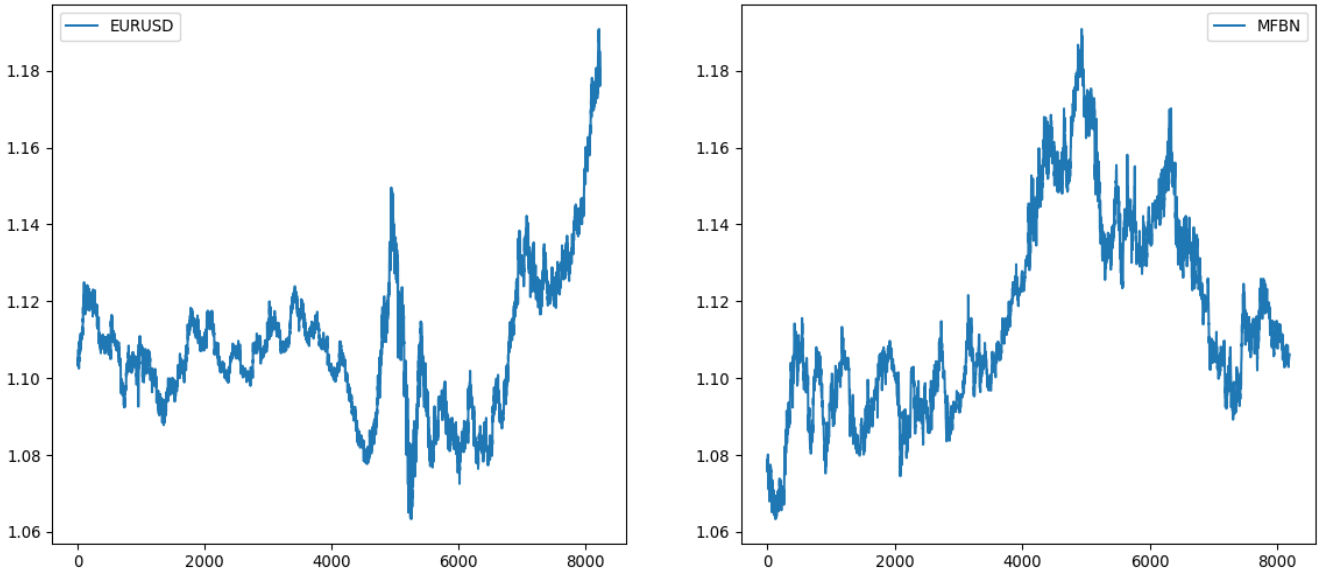

Рисунок 2.10 – Графік годинних котирувань (зліва), та графік мультифрактального броунівського руху побудованого по фрактальним властивостей котирувань (справа)

ВИСНОВКИ ПО ДРУГОМУ РОЗДІЛУ

1.В практичних задачах виникає проблема недостатнього об’єму даних для навчання моделі машинного навчання, що створює перешкоду для використання ефективних методів розв’язання задачі. В таких ситуаціях використовують методи математичного моделювання, котрі дозволяють згенерувати необхідну кількість навчальних прикладів з заданими властивостями. У роботі здійснено моделювання масивів даних з різними фрактальними властивостями, таких як: фрактальний броунівський рух, експоненціальне перетворення фрактального гаусівського шуму, стохастичний біноміальній мультиплікативний каскад, мультифрактальний броунівській рух.

2. Отримав подальший розвиток метод моделювання мультифрактальних УМД, а саме стохастичного мультиплікативного біноміального каскаду. В рамках запропонованого методу вирішена задача нелінійної оптимізації, що дозволило генерувати стохастичний мультиплікативний біноміальний каскад з більш різноманітними мультифрактальними властивостями.

3. Використання запропонованого підходу дає змогу моделювати мультифрактальних броунівський рух на основі оцінок узагальненого показника Херста , отриманих за реальними даними.

4.Таким чином у роботі отримали подальший розвиток методи генерації фрактальних даних, що дозволило охопити практично весь діапазон мультифрактальних властивостей упорядкованих фрактальних даних, які зустрічаються у практичних задачах.

Результати другого розділу опубліковані в [20, 21, 22, 61, 90, 107, 117, 118].

РОЗДІЛ 3
МЕТОДИ МАШИННОГО НАВЧАННЯ ЯК ІНСТРУМЕНТ ДЛЯ ФУНКЦІОНУВАННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ КЛАСИФІКАЦІЇ УПОРЯДКОВАНИХ МАСИВІВ ДАНИХ

Для задоволення вимоги до ІТ про оптимальність класифікації різних станів досліджуваного об’єкту необхідно визначити залежність точності класифікації методів машинного навчання від характеристик різних типів УМД.

У розділі представлено опис та результати чисельних експериментів класифікації УМД в залежності від їх типу, 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 скінчена множина УМД, для яких відомо, до яких класів вони належать. Ця множина є навчальною вибіркою. Класова приналежність інших УМД не відома. Необхідно побудувати алгоритм, здатний класифікувати довільний УМД з вихідної множини. Було розглянуто існуючі підходи до класифікації УМД різної природи [9, 13, 14, 23, 42, 63, 64].

В формальному виді постановка задачі класифікації має такий вигляд. Нехай $x_i \in X, i = \overline{1,n}$ - множина об’єктів 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 розмічений об’єкт, або прецедент. Кінцева множина $\{x_i\}, i = \overline{1,n}$ представляє собою матрицю $\{x_{i,j}\}, i = \overline{1,\mathrm{n}}, j = \overline{1,m}$ розміром $n \times m$, де рядок матриці – це масив ознак одного об’єкта, $\{y_i\}, i = \overline{1,n}$ - вектор відповідей, елемент якого є значення номеру класу. Комбінація $\{x_i\}, i = \overline{1,n}$ та $\{y_i\}, i = \overline{1,n}$ називається навчальною вибіркою. Задача класифікації полягає у визначенні функції залежності $f : X \to Y$ котра пророкує 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 і ряду статистичних властивостей, які визначаються специфікою даного УМД. Таким чином, кажучи про деякий клас УМД з близькими фрактальними властивостями, ми 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.

3.1.Розбиття на класи

Експерименти проводилися двох типів, а саме: визначення значення показника Херста методами машинного навчання, та визначення наявності ДДоС атаки у часовому ряді базуючись на мультифрактальних властивостях.

Для задачі визначення показника Херста експерименти проводилися на декількох конфігураціях вихідних даних моделей. Мінімальним та максимальним значенням показника Херста для персистентніх часових рядів були рівні 0,51 та 0,99 відповідно. Для класифікації часових рядів на всьому інтервалі показника Херста максимальним значеннями було 0,99, а мінімальним 0,01. Повний список конфігурації параметрів показника Херста для різних експериментів можна знайти в таблиці 3.1.

Для визначення присутності ДДоС атаки в модельних даних використовувалась бінарна класифікація, оскільки класів було всього два, а саме наявність атаки в часовому ряді, моделювання мережевої активності, та відсутність такої. Значення показника Херста змінювалось в інтервалі від 0,5 до 1,0 з кроком 0,1, при цьому мінімальне та максимальне значення були 0,51 та 0,99 відповідно. Величина показника Херста для моделювання часового ряду бралася випадково по рівномірному закону розподілу у відповідному інтервалі значень показника Херста.

Для визначення належності часового ряду до одного з класів були використані декілька з найефективніших методів машинного навчання, а саме беггінг, випадковий ліс, багатошаровий персептрон, нейронні мережі довгої

короткострокової пам’яті. В перших двох методах були використані дерева прийняття рішень як класифікації так і регресії. Навчання моделей для кожного класу проводилось на 300 прикладах часових рядів навчання та перевірялось на 150 тестових.

Таблиця 3.1 – Співвідношення діапазону показника Херста, ширини інтервалу та кількості інтервалів класифікації показника Херста

№	$H_{\min}$	$H_{\max}$	H_{step}	Кількість класів	Тип задачі класифікації
1	0,51	0,99	0,05	10	Значення показника Херста
2	0,51	0,99	0,1	5	Значення показника Херста
3	0,01	0,49	0,05	10	Значення показника Херста
4	0,01	0,49	0,1	5	Значення показника Херста
5	0,51	0,99	0,1	2	Визначення наявності ДДоС атаки

3.2. Класифікатори

В якості методів машинного навчання використовувалися наступні алгоритми: метод беггінг з деревами класифікації та регресії[18]; метод випадкового лісу з деревами класифікації та регресії[19]; повнозв’язні нейронні мережі, на основі багатошарового персептрону та нейронні мережі довгої короткострокової пам'яті (Long Short Term Memory, LSTM) - особливий вид рекурентних нейронних мереж, здатних до навчання довгостроковим залежностям[68, 81].

Беггінг

Беггінг – технологія класифікації, де всі елементарні класифікатори навчаються та працюють незалежно один від одного. Ідея полягає в тому, що класифікатори не виправляють помилки друг друга, а компенсують їх при голосуванні. Для отримання результату роботи ансамблю моделей зазвичай використовують наступні способи комбінації: голосування або усереднення. Ефективність беггінг 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 попасти до деяких навчальних підвибірок [27, 36].

Випадковий ліс

Випадковий ліс, також є методом беггінг, але на відміну 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 здійснюється випадком чином ознак; для кожної підвибірки дерево рішень будується до повного вичерпання навчальних прикладів та не піддається процедурі відсічення гілок.

Нейронні мережі

Для проведення класифікації в якості нейронних мереж був вибраний багатошаровий персептрон з функцією активації типу реле [27, 36]. Ця нейронна мережа є універсальним апроксиматором та здатна виявляти приховані закономірності в даних [65]. При правильному підборі архітектури нейронної мережі навчена модель здатна давати достатньо добрі результати класифікації. Для попередження перенавчання моделі та для збільшення точності класифікації в структуру нейронної мережі були включені декілька шарів пакетної нормалізації [48].

LSTM

Основною її відмінністю мережи LSTM від звичайної мережі зі зворотними зв’язками є те, що кожний нейрон прихованого шару LSTM мережі замінений на комірку пам’яті [2, 68, 81]. 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 повнозв’язними шарами персептронів з функцією активації типу ReLu та шаром пакетної нормалізації.

3.3. Характеристики якості класифікації для порівняння

Матриця помилок та коефіцієнт точності

Як відомо, метою і змістом задачі класифікації є віднесення об'єкту до відповідного класу (категорії). Позначимо через показник x_{ij} кількість об'єктів, які відкласифіковані як такі, що належать до категорії C_j хоча в дійсності вони належать до категорії $C_i, i, j = 1,2,...,r$, $r-$ загальна кількість класів (категорій) об'єктів. З таких показників формується матриця помилок

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{r1} & \cdots & x_{rr} \end{pmatrix},$$

яка є інформативною основою для оцінювання точності класифікації.

Надалі будемо використовувати такі позначення:

$\sum_{j=1}^{r} x_{ij} \equiv x_{i+}$ – кількість об'єктів n_i з категорії C_i;

$\sum_{i=1}^{r} x_{ij} \equiv x_{+j}$ – кількість об'єктів, віднесених при класифікації до категорії C_j;

$\sum_{i=1}^{r} n_i \equiv N$ – кількість усіх ділянок.

З матриці помилок отримуються коефіцієнти для оцінки точності, а саме:

Середньо класова точність

$$C(X) = \frac{1}{r}\sum_{i=1}^{r} \frac{x_{ii}}{n_i},$$

Загальна точність класифікації

$$accuracy = \frac{1}{N}\sum_{i=1}^{r} x_{ii},$$

Precision, recall та F_β -міра

Для оцінки якості роботи алгоритму на кожному з класів введені метрики precision(точність) та recall (повнота):

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

де TP - кількість об'єктів котрі є позитивними алгоритм відніс до позитивних, FP - похибка першого роду, FN - похибка другого роду.

Precision можна пояснити як долю об'єктів названих класифікатором позитивними та при цьому дійсно є позитивними, а recall показує, яку частку об'єктів позитивного класу з усіх об'єктів позитивного класу знайшов алгоритм. Precision та recall не залежить, на відміну від accuracy, від співвідношення класів і тому використовується в умовах незбалансованих вибірках.

Існує декілька способів об'єднати precision та recall в агрегований критерій якості. F_β - міра – середнє гармонічне precision та recall

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{precision \cdot recall}{\left(\beta^2 \cdot precision\right) + recall},$$

β в даному випадку визначає вагу точності в метриці, та при $\beta = 1$ це середнє гармонічне.

F_β – міра набуває максимального значення коли повнота та точність, рівні одиниці, та близька до нуля, якщо один з аргументів близький до нуля.

3.4.Ознаки

3.4.1. Ознаки – значення часового ряду

Класифікація проводилась для часових рядів з різноманітними мультифрактальними властивостями, отриманих генерацією різноманітними моделями мультифрактальних часових рядів.

На рисунку 3.1 типові каскади з різних класів: ліворуч каскад з показником Херста $H = 0,72$, отриманий за допомогою симетричного бета-розподілу з параметром $\alpha = 0,9$, праворуч каскад, для якого $H = 0,9$ та $\alpha = 3$.

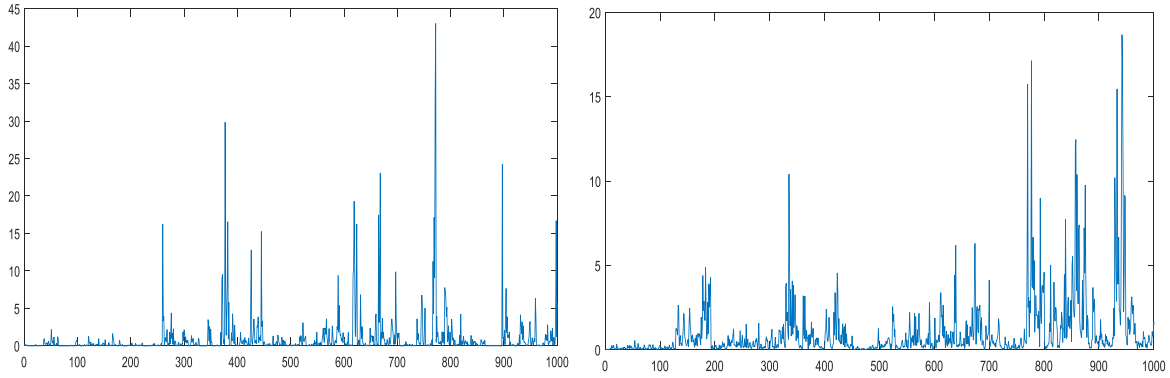

Рисунок 3.1 – Мультифрактальні каскади: ліворуч $H = 0,72$, праворуч $H = 0,9$

В даному випадку кожний клас представляє собою набір згенерованих часових рядів з однаковим показником Хреста. У випадку, коли вагові коефіцієнти є значеннями симетричного бета-розподілу, скейлінгова експонента $\tau(q)$ однозначно визначається значенням показника Херста H, $0,5 < H < 1$. Значення показника Херста змінювалось в інтервалі від 0,5 до 1,0 з кроком 0,05, при цьому мінімальне та максимальне значення були 0,51 та 0,99 відповідно, таким чином навчання моделі проводилось на 10 класах.

Для визначення належності часового ряду до одного з 10 класів були використані декілька з найефективніших методів машинного навчання, а саме беггінг, випадковий ліс, багатошаровий персептрон, нейронні мережі довгої короткострокової пам’яті. В перших двох методах були використані дерева прийняття рішень як класифікації так і регресії. Навчання моделей для кожного класу проводилось на 1000 прикладах часових рядів навчання та перевірялось на 250 тестових.

На рисунку 3.2 представлені результати роботи моделі на навчальній вибірці з 11 000 прикладів, верхній графік, та нижній графік – 2750 прикладів тестової вибірки.

На рисунку 3.3 представлені гістограми розподілення ймовірності передбачення номеру класу для кожного значення показника Херста, отримані методом випадковий ліс з деревами регресії, та відповідають результатам на рисунку 3.2. Значення H показано над кожною гістограмою. Очевидно, що каскади, побудовані для екстремальних значень показника Херста $H = 0{,}51;\ 0{,}99$ визначаються практично в 100% випадках. Це пояснюється їх сильно вираженими мультифрактальними властивостями.

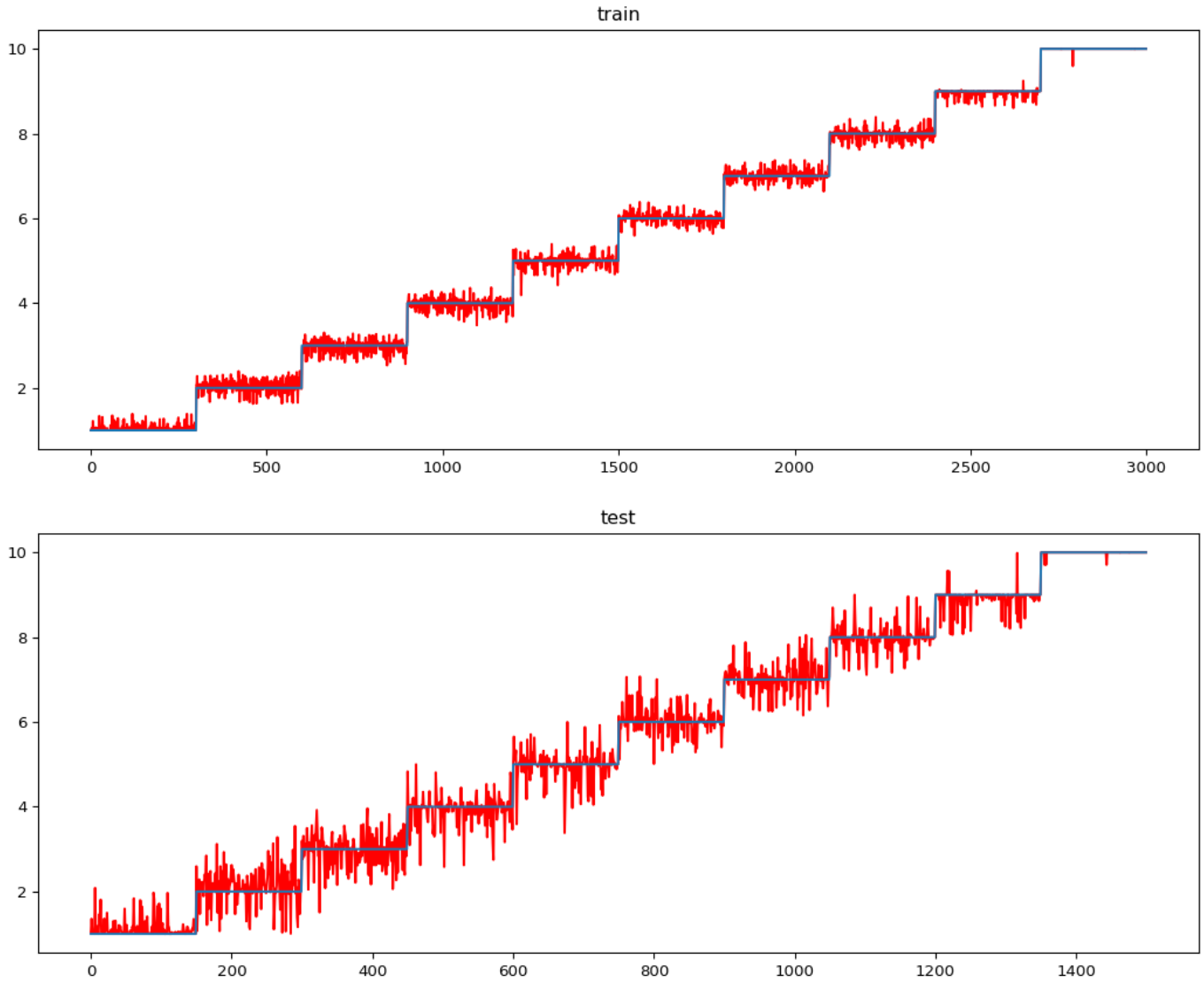

Рисунок 3.2 – Результати класифікації: верхній графік – навчальна вибірка, нижній – тестова

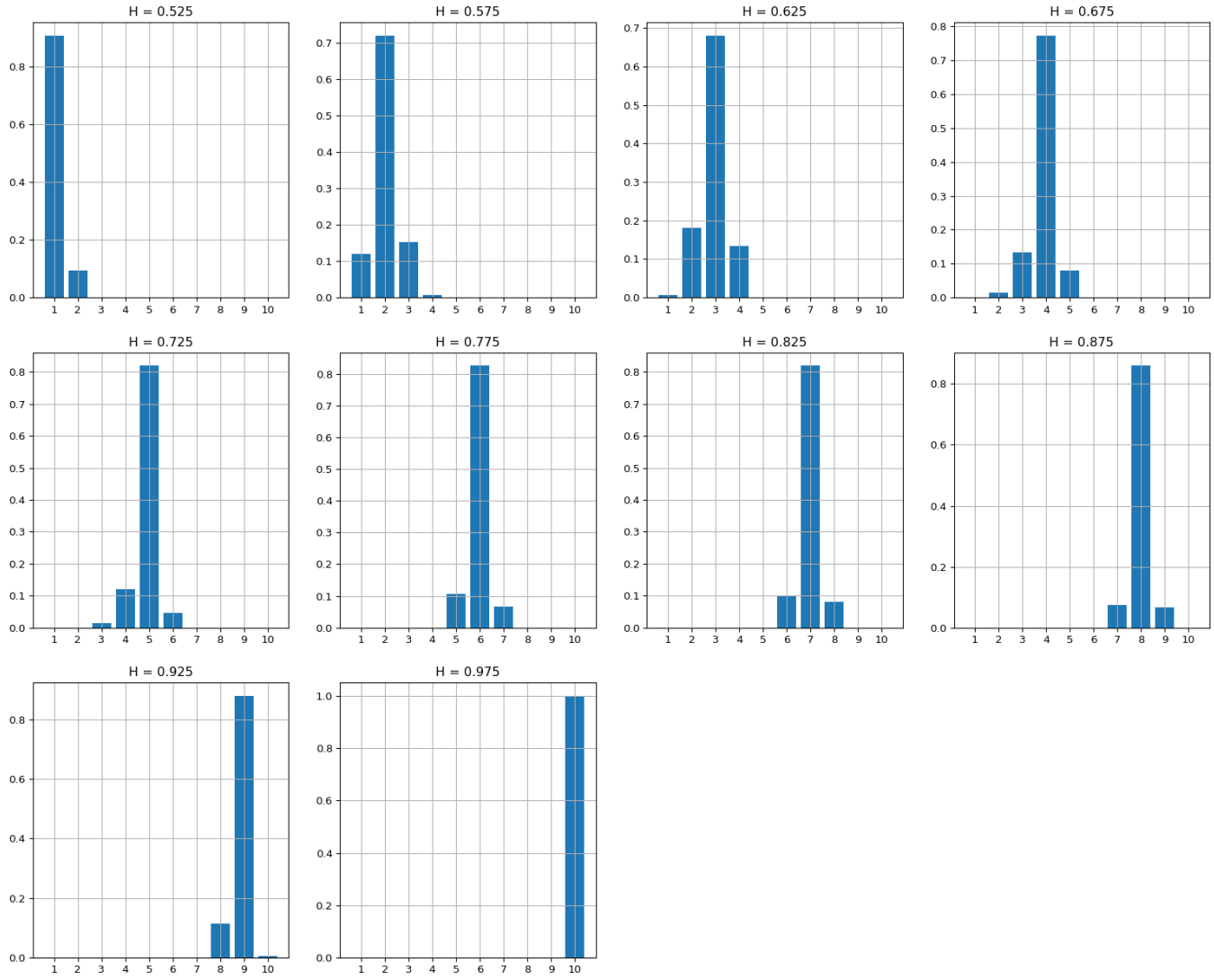

Рисунок 3.3 – Розподілення ймовірностей визначення номеру класу в залежності від значення показника Херста для мультиплікативних каскадів та випадкового лісу з деревами регресії

В таблицях 3.2–3.8 представлені середні ймовірності визначення класу в залежності від довжини мультифрактальних рядів, моделі генерації часового ряду та метода машинного навчання. Результати показують, що використовувати методи машинного навчання для класифікації часових рядів по даним самих часових рядів можна використовувати лише для рядів біноміальний мультиплікативний каскад та рядів типу експоненціальний гаусівський шум, при чому для антиперсистених рядів гаусівського шуму точність класифікації дещо більша. При цьому використання методу Випадковий ліс дає більш кращі

результати порівняно з методом Беггінг для кожного типу дерев прийняття рішень, але залежності від типу дерев не проявляється. Використання нейронних мереж, особливо LSTM не дають добрих результатів, основною причиною цього є той факт що у часовому ряді з невеликим коефіцієнтом Херста слабо проявляються детерміністичні залежності.

Таблиця 3.2 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для персистентного діапазону значень показника Херста $H > 0{,}5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,138	0,117	0,159	0,111	0,171	0,111

Таблиця 3.3 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,530	0,438	0,577	0,441	0,259	0,121
1024	0,567	0,498	0,648	0,491	0,292	0,108
2048	0,610	0,521	0,653	0,524	0,351	0,109
4096	0,665	0,548	0,549	0,548	0,360	0,100

Таблиця 3.4 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,172	0,111	0,16	0,115	0,131	0,131

Таблиця 3.5 – Середня ймовірність передбачення класу для часових рядів типу біноміальний мультиплікативний каскад в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина СБМК	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,821	0,833	0,849	0,835	0,411	0,411
1024	0,86	0,879	0,891	0,875	0,417	0,415
2048	0,897	0,916	0,905	0,912	0,485	0,485
4096	0,916	0,935	0,939	0,939	0,517	0,515

Таблиця 3.6 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для антиперсистентного діапазону значень показника Херста $H < 0,5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,261	0,198	0,29	0,192	0,201	0,191

Таблиця 3.7 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,600	0,549	0,642	0,541	0,293	0,318
1024	0,676	0,582	0,772	0,580	0,371	0,377
2048	0,733	0,631	0,784	0,615	0,437	0,454
4096	0,787	0,667	0,843	0,668	0,562	0,571

Таблиця 3.8 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,261	0,177	0,268	0,183	0,214	0,198

3.4.2. Ознаки – статистичні та мультифрактальні характеристики

В таблицях 3.9–3.15 представлені середні ймовірності визначення класу в залежності від довжини мультифрактальних рядів, моделі генерації часового ряду та метода машинного навчання. Результати моделювання показують, що класифікація часових рядів по фрактальним та статистичним характеристикам дає точніші результати ніж класифікація по даним самих часових, але точність все ще недостатня щоб використовувати даний підхід для класифікації різноманітних ЧР. Точність стала дещо краща для рядів типу експоненціальний гаусівський шум та біноміальний мультиплікативний каскад, але занадто мала для фрактального гаусівського

Таблиця 3.9 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для персистентного діапазону значень показника Херста $H > 0,5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,154	0,156	0,156	0,155	0,175	0,160

Таблиця 3.10 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,918	0,921	0,918	0,921	0,911	0,915
1024	0,969	0,968	0,967	0,969	0,943	0,949
2048	0,983	0,983	0,986	0,983	0,958	0,965
4096	0,992	0,993	0,993	0,992	0,967	0,979

Таблиця 3.11 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,166	0,141	0,167	0,127	0,155	0,149

Таблиця 3.12 – Середня ймовірність передбачення класу для часових рядів типу біноміальний мультиплікативний каскад в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина БМК	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,803	0,804	0,803	0,805	0,705	0,711
1024	0,864	0,866	0,865	0,867	0,755	0,781
2048	0,881	0,885	0,875	0,883	0,795	0,814
4096	0,919	0,919	0,917	0,92	0,811	0,849

Таблиця 3.13 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для антиперсистентного діапазону значень показника Херста $H < 0,5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,312	0,285	0,299	0,281	0,318	0,317
1024	0,365	0,347	0,378	0,349	0,349	0,353
2048	0,437	0,416	0,434	0,411	0,441	0,419
4096	0,497	0,503	0,479	0,506	0,485	0,495

Таблиця 3.14 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,979	0,979	0,979	0,979	0,948	0,979
1024	0,984	0,983	0,984	0,983	0,970	0,983
2048	0,987	0,988	0,987	0,986	0,975	0,986
4096	0,992	0,992	0,992	0,992	0,980	0,991

Таблиця 3.15 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,385	0,376	0,384	0,377	0,363	0,368
1024	0,414	0,413	0,421	0,413	0,429	0,433
2048	0,498	0,491	0,499	0,495	0,512	0,518
4096	0,532	0,517	0,533	0,518	0,552	0,561

3.4.3. Ознаки – фрактальні та рекурентні характеристики

Головна ідея застосування методів нелінійної динаміки до аналізу траєкторії динамічної системи полягає в тому що основна структура, яка містить в собі всю інформацію про систему, а саме, атрактор системи, може бути відновлена через вимір однієї компоненти цієї системи [74]. Широко використовувана процедура Паккарда-Такенса дозволяє відновити фазову траєкторію динамічної системи по одній реалізації:

$$F(t)=[x(t),x(t+\tau),...,x(t+m\tau)],$$

де: $F(t)$ – m- мірний псевдофазовий простір, $x(t)$ – часова реалізація системи, τ – період запізнення.

Рекурентна діаграма є проекцією m- мірного псевдофазового простору на площину [75]. Нехай точка x_i відповідає точці фазової траєкторії $x(t)$, котра описує динамічно систему в m-мірному просторі в момент часу $t=i$, для $i=1,...,N$, тоді рекурентна діаграма RP є масив точок, де ненульовий елемент з координатами (i,j) відповідає випадку коли відстань між x_j та x_i є меншою ніж ε

$$RP_{i,j}=\Theta(\varepsilon-\| x_i-x_j \|),\ \ x_i,x_J\in R^m,\ i,j=1,...N,$$

де ε – розмір околиці точки x_i, $\left\|x_i-x_j\right\|$ – відстань між точками, $\Theta(\cdot)$ – відстань між точками.

Аналіз топології діаграми дозволяє класифікувати спостерігаючи процеси: визначати однорідні процеси з незалежними випадковими значеннями; процеси з повільно змінюваними параметрами; періодичні та осцилюючі процеси, котрі відповідають нелінійним системам і т. д. Чисельний аналіз рекурентних діаграм дозволяє обчислювати міри складності структур рекурентних діаграм, такі як міра рекурентності, міра детермінізму, міра ентропії та інші.

Міра рекурентності (recurrence rate, RR)

$$RR=\frac{1}{N^2}\sum\nolimits_{i,j=1}^{N}RP_{i,j}^{\varepsilon}$$

показує густину рекурентних точок, просто підраховуючи їх. Дана міра показує ймовірність находження рекурентної точки в рекурентній діаграмі. (ймовірність повторення стану).

Наступна міра розглядає діагональні лінії. Частотне розподілення довжин l діагональної ліній в RP $P^{\varepsilon}(l)=\{l_l;i=1...N_l\}$, де l_i – довжина i- ї діагональної лінії, N_l- кількість діагональних ліній (кожна лінія рахується лише один раз). Стохастичні часові ряди можуть породжувати дуже короткі діагоналі, або зовсім

не породжувати їх, в той час як детерміністичні процеси дають довгі діагоналі та малу кількість окремих рекурентних точок. Таким чином, співвідношення рекурентних точок

$$DET = \frac{\sum_{l=l_{\min}}^{N} l\, P^{\varepsilon}(l)}{\sum_{i,j}^{N} R_{i,j}^{\varepsilon}(l)},$$

називається мірою детермінізму (determinism, DET) або прогнозованість системи. Варто відзначити, що ця міра не володіє знаннями реального детермінізму процесу. При $l_{\min} = 1$ виконується співвідношення $DET = RR$.

Діагональні структури показують час, протягом якого ділянка траєкторії підходить достатньо близько до другої ділянки траєкторії. Таким чином, ці лінії дозволяють судити про розходження елементів траєкторії. Середня довжина діагональних ліній

$$L = \frac{\sum_{l=l_{\min}}^{N} lP^{\varepsilon}(l)}{\sum_{l=l_{\min}}^{N} P^{\varepsilon}(l)},$$

де L - це середній час, протягом якого дві ділянки траєкторії проходять поруч один до одного, та може розглядатися як середній час прогнозованості.

Міра ентропії (entropy, $ENTR$) співвідноситься з ентропією Шеннона (Shannon) частотного розподілу дожин діагональних ліній:

$$ENTR = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l),$$

$$p(l) = \frac{P^{\varepsilon}(l)}{\sum_{l=l_{\min}}^{N} P^{\varepsilon}(l)},$$

та відображає складність детерміністської складової в системі.

Розглянемо точку траєкторії x_i та множину зв'язаних з нею рекурентних точок, котрі формують вертикальні структури на RP. Позначимо $P_i(\nu) = \{\nu_k; k = 1,2,...,K\}$ множину довжин вертикальних структур та визначимо розподіл довжин вертикальних ліній для всього RP в цілому.

Міра завмирання (laminarity, LAM)

$$LAM = \frac{\sum_{\nu=\nu_{\min}}^{\tilde{N}} \nu P^{\varepsilon}(\nu)}{\sum_{i,j}^{N} RP_{i,j}^{\varepsilon}},$$

визначається відношенням кількості рекурентних точок, котрі утворюють вертикальні лінії, до загальної кількості рекурентних точок. Величина LAM характеризує наявність станів завмирання системи (тобто коли рух системи по фазовій траєкторії зупиняється або є дуже повільним). Середня довжина вертикальних структур

$$TT = \frac{\sum_{\nu}^{\tilde{N}} \nu P^{\varepsilon}(\nu)}{\sum_{i,j}^{N} P^{\varepsilon}(\nu)},$$

називається мірою часу зупинки(trapping time, TT) та характеризує середній час, який система може провести в околиці деякого стану.

В роботах [Запорожье, ИТЕА] показано, що фрактальним ЧР з різним показником Херста відповідають різні рекурентні структури та відповідні кількісні характеристики.

На рисунку 3.4, 3.5 та в таблиці 3.15 приведені результати рекурентного аналізу реалізацій з різними показниками Херста. На рисунку 3.4 приведені реалізації та діаграми фрактального броунівського руху з показником Херста $H = 0.6$ (зверху) та $H = 0.9$ (знизу). Реалізації з більшим параметром Херста більш персистентні, вони мають набагато меншу порізаність. На рекурентній діаграмі реалізації з більшою персистентністю шикуються вздовж діагональної лінії, в той час як для рядів з невеликим показником Херста рекурентні точки заповнюють більшу область фазового простору.

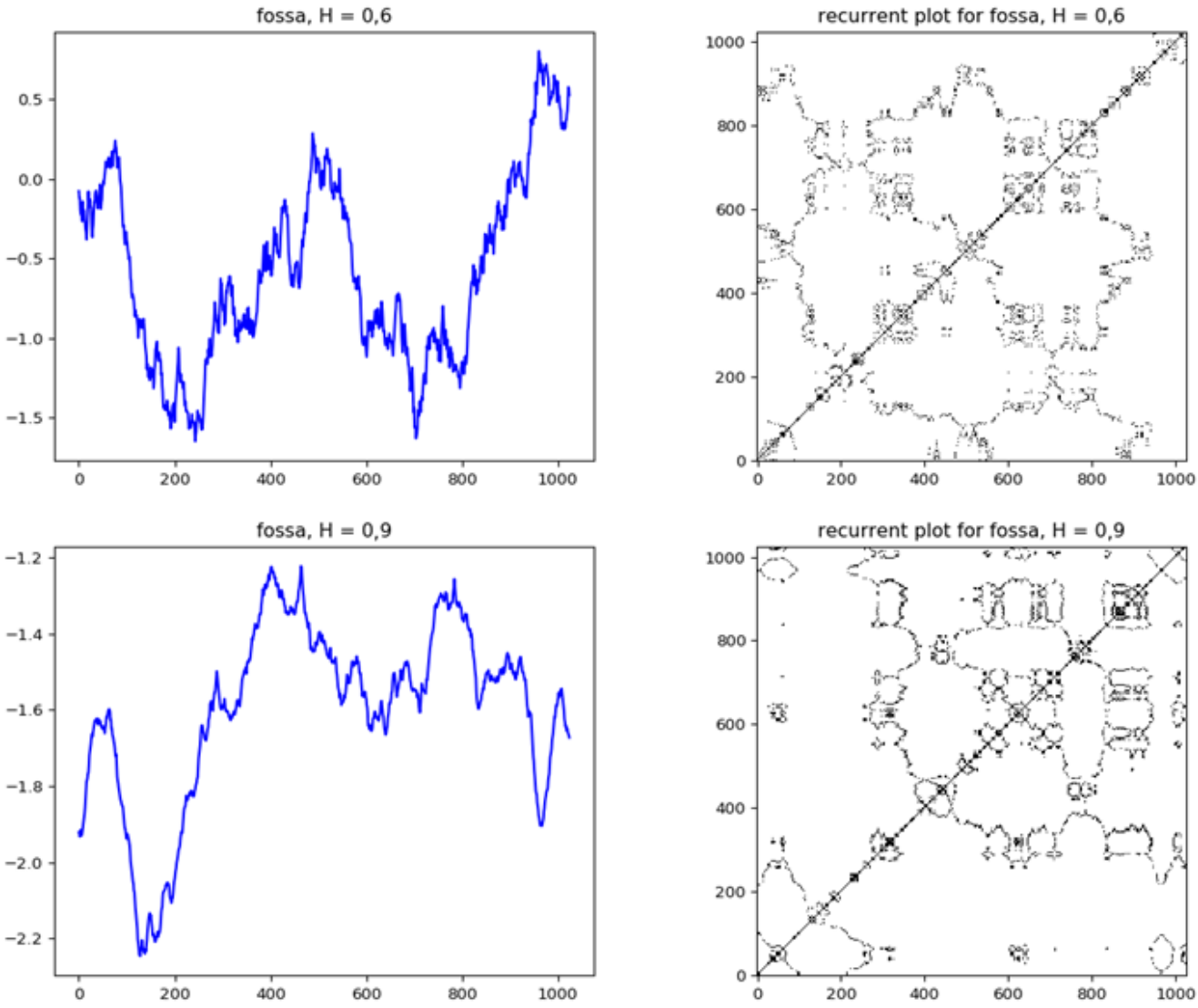

Рисунок 3.4 – Реалізації рекурентних діаграм для фрактального броунівського руху при $H = 0,6$ (зверху) та $H = 0,9$(знизу).

На рисунку 3.5 приведені відповідні реалізації РД для фрактального гаусівського шуму з показником Херста $H = 0,6$ (зверху) та $H = 0,9$ (знизу). При зростанні показника Хреста з’являється нерівномірна структурованість фазового простору, яка відсутня для рядів з параметром Херста близьким до 0,5.

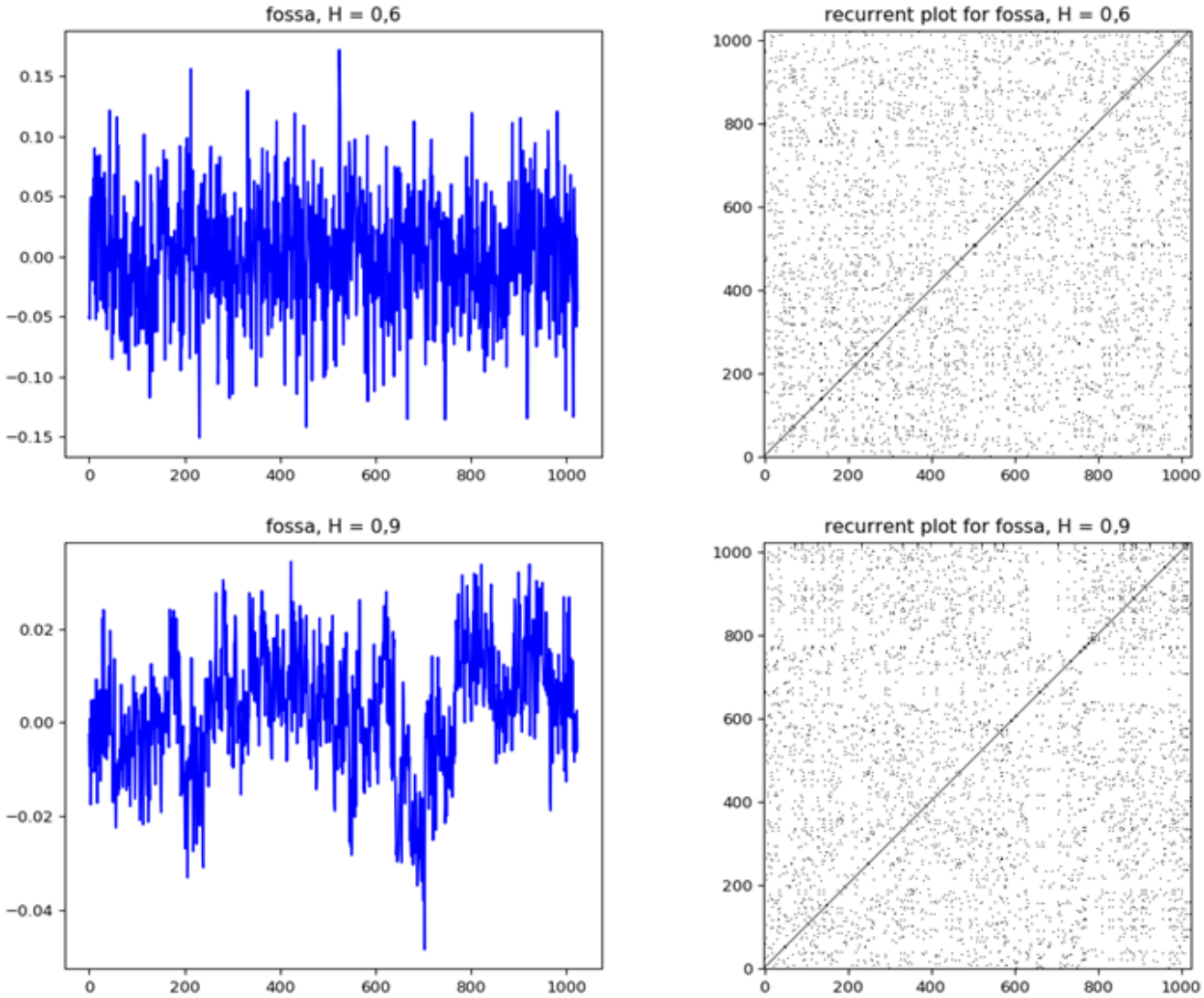

Рисунок 3.5 – Реалізації РД для ФГШ при $H = 0{,}6$ (зверху) та $H = 0{,}9$ (знизу).

Структурні відмінності знайшли своє відображення в кількістних характеристиках. В таблиці 3.16 представлені середні значения характеристик РД котрі демонструють, що зі збільшенням показника Херста збільшуються детерміністичні складові, а отже прогнозованність ЧР. Цікаво відзначити, що хоча густина точок (RR) у реалізації ФБР на порядок нижче, ніж у ФГШ, середня довжина діагональних ліній (L) та середній час, який система може провести в деякому стані (TT) в декілька разів більший. Міра детермінізма (Det) в обох випадках близька до показника Херста H.

Таблиця 3.16 – Значення рекурентних характеристик для деяких часових рядів

	H			
	ФБД		ФГШ	
	0,6	0,9	0,6	0.9
RR	0,027	0,049	0,188	0,228
Det	0,684	0,937	0,615	0,889
L	3,866	23,292	2,763	3,180
ENTR	1,826	3,831	1,206	1,508
LAM	0,824	0,937	0,476	0,717
TT	4,010	22,235	2,659	15,197

Таким чином кількісні характеристики РД можуть бути використані в якості ознак для проведення класифікації за допомогою машинного навчання.

Для проведення чисельного експерименту була згенерована множина часових рядів ФБР та відповідних ФГШ з різноманітними показниками Херста. Кожен клас ЧР представляє собою множину часових рядів з показником Херста, значення якого належить заданому діапазону. Значення діапазонів показника Херста змінювалось в інтервалі від 0,51 до 0,99 з кроком 0,05 Таким чином навчання моделей виконувалось на 10 класах, де $H \in \{[0,51;0,55),...,[0,95;0,99]\}$. В середині кожного класу значення H обиралось за допомогою рівномірного закону розподілу ймовірностей.

Для визначення належності часового ряду до одного з 10 класів були використані метод випадкового лісу, метод беггінг з деревами регресії та класифікації, класичні нейронні мережі, а також рекурентні нейронні мережі.

Навчання моделей для кожного класу проводилось на 300 прикладах часових рядів навчання та перевірялись на 150-и тестових. Класифікація проводилась для часових рядів різної довжини.

В даній роботі запропоновано використовувати рекурентні характеристики представлені в розділі 3.4.2. Для кожного УМД будувалась рекурентна діаграма та

обчислювались відповідні рекурентні характеристики, на базі яких проводилось навчання и класифікація рядів методами машинного навчання.

В числових дослідженнях окремо були розглянуті ряди ФБР та ФГШ. Окрім цього класифікація проводилась як з використанням показника Херста в якості додаткової ознаки, так і без нього. В ході проведення дослідження було відзначено, що результати класифікації персистентних так антиперсистентних рядів мають вагомі відмінності та були проведені дослідження для обох випадків.

На мал. 6 представлені гістограми розподілення ймовірностей визначення номеру класу для кожного діапазону значень показника Херста, отримані на основі рекурентних характеристик для рядів ФБР з урахуванням значення показника Херста *H* як одного з ознак. Над кожною гістограмою показано середня значення діапазону *H*.

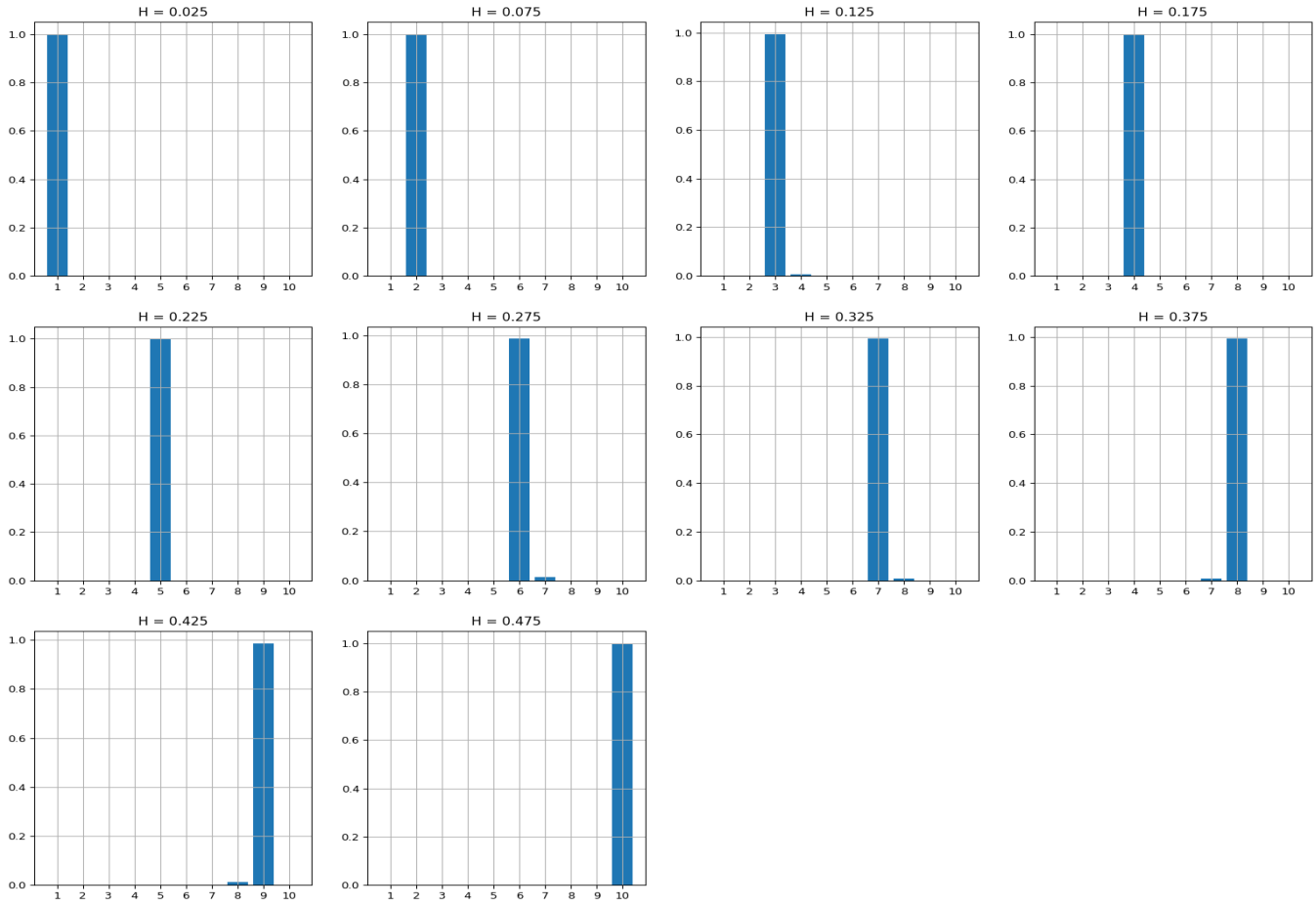

Рисунок 3.6 – Розподілення ймовірностей визначення номеру класу в залежності від діапазону *H*

В таблицях 3.17–3.23 представлені середні ймовірності визначення класу в залежності від довжини мультифрактальних рядів, моделі генерації часового ряду та метода машинного навчання. Результати дослідження показують що для класифікації фрактального та мультифрактального броунівського шуму можна використовувати нейронні мережі які дають прийнятну точність класифікації.

Таблиця 3.17 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для персистентного діапазону значень показника Херста $H > 0,5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,883	0,891	0,773	0,888	0,751	0,906
1024	0,943	0,942	0,855	0,915	0,769	0,964
2048	0,971	0,970	0,900	0,971	0,809	0,982
4096	0,992	0,993	0,991	0,991	0,849	0,991

Таблиця 3.18 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,934	0,937	0,941	0,943	0,465	0,414
1024	0,985	0,987	0,985	0,988	0,527	0,489
2048	0,991	0,994	0,991	0,994	0,621	0,543
4096	0,997	0,997	0,997	0,997	0,654	0,581

Таблиця 3.19 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,547	0,564	0,475	0,562	0,587	0,594
1024	0,781	0,796	0,631	0,685	0,627	0,785
2048	0,918	0,936	0,769	0,933	0,698	0,940
4096	0,956	0,961	0,957	0,962	0,764	0,969

Таблиця 3.20 – Середня ймовірність передбачення класу для часових рядів типу біноміальний мультиплікативний каскад в залежності від довжини ряду та метода машинного навчання. Персистентний діапазон значень показника Херста

Довжина БМК	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,869	0,88	0,855	0,877	0,625	0,605
1024	0,921	0,92	0,902	0,921	0,684	0,652
2048	0,938	0,938	0,941	0,937	0,698	0,661
4096	0,952	0,955	0,955	0,954	0,709	0,674

Таблиця 3.21 – Середня ймовірність передбачення класу для часових рядів типу фрактальний броунівський рух для антиперсистентного діапазону значень показника Херста $H < 0,5$, в залежності від довжини ряду та метода машинного навчання

Довжина ФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,839	0,832	0,785	0,833	0,602	0,905
1024	0,975	0,973	0,931	0,974	0,647	0,953
2048	0,997	0,997	0,965	0,997	0,684	0,961
4096	0,996	0,995	0,97	0,995	0,727	0,969

Таблиця 3.22 – Середня ймовірність передбачення класу для часових рядів типу експоненціальний гаусівський шум в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина ЕГШ	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,980	0,981	0,980	0,980	0,952	0,980
1024	0,982	0,985	0,982	0,984	0,971	0,981
2048	0,988	0,987	0,988	0,986	0,976	0,989
4096	0,992	0,992	0,992	0,992	0,980	0,991

Таблиця 3.23 – Середня ймовірність передбачення класу для часових рядів типу мультифрактальний броунівський рух в залежності від довжини ряду та метода машинного навчання. Антиперсистентний діапазон значень показника Херста

Довжина МФБР	Беггінг		Випадковий ліс		Багатошаровий персептрон	LSTM
	Дерева класифікації	Дерева регресії	Дерева класифікації	Дерева регресії		
512	0,582	0,561	0,559	0,557	0,580	0,601
1024	0,611	0,585	0,614	0,583	0,615	0,637
2048	0,665	0,647	0,67	0,645	0,672	0,679
4096	0,726	0,7	0,707	0,698	0,735	0,741

3.5 Оцінювання показника Херста методами машинного навчання

Загальноприйнятим підходом до класифікації фрактальних упорядкованих даних є оцінювання показника Херста. Одним з популярних і досить точним методом оцінювання є метод мультифрактального детрендованого флуктуаційного аналізу.

Було проведено велика кількість експериментів по оцінюванню показника Херста та його довірчого інтервалу по значенням УМД як методом МФДФА так і методами машинного навчання. В ході експериментів використовувалися дані модельних рядів таких як фрактальний, мультифрактальних броунівський рух, стохастичний мультиплікативний біноміальний каскад, експоненціальне перетворення фрактального гаусівського шуму, при малих довжинах рядів, а саме 256, 512 та 1024 значення. При використанні методів машинного навчання в якості ознак використовувалися значення УМД, статистичні та фрактальні характеристики та характеристики рекурентних діаграм.

Найбільш точні результати дають статистичні та фрактальні характеристики, а також рекурентні характеристики використання яких залежить від мультифрактальних характеристик досліджуваних УМД.

На рисунках 3.7 та 3.8 представлені результати визначення показника Херста методом МФДФА, та за допомогу нейронних мереж довгої короткострокової пам’яті з використанням в якості ознак статистичних та фрактальних характеристик.

В таблиці 3.24 представлені статистичні оцінки ширини довічого інтервалу для різних типів УМД при оцінюванні методом МФДФА та методами машинного навчання, з таблиці видо що довірчий інтервал більше ніж у 2 рази вужчій при оцінюванні методами машинного навчання.

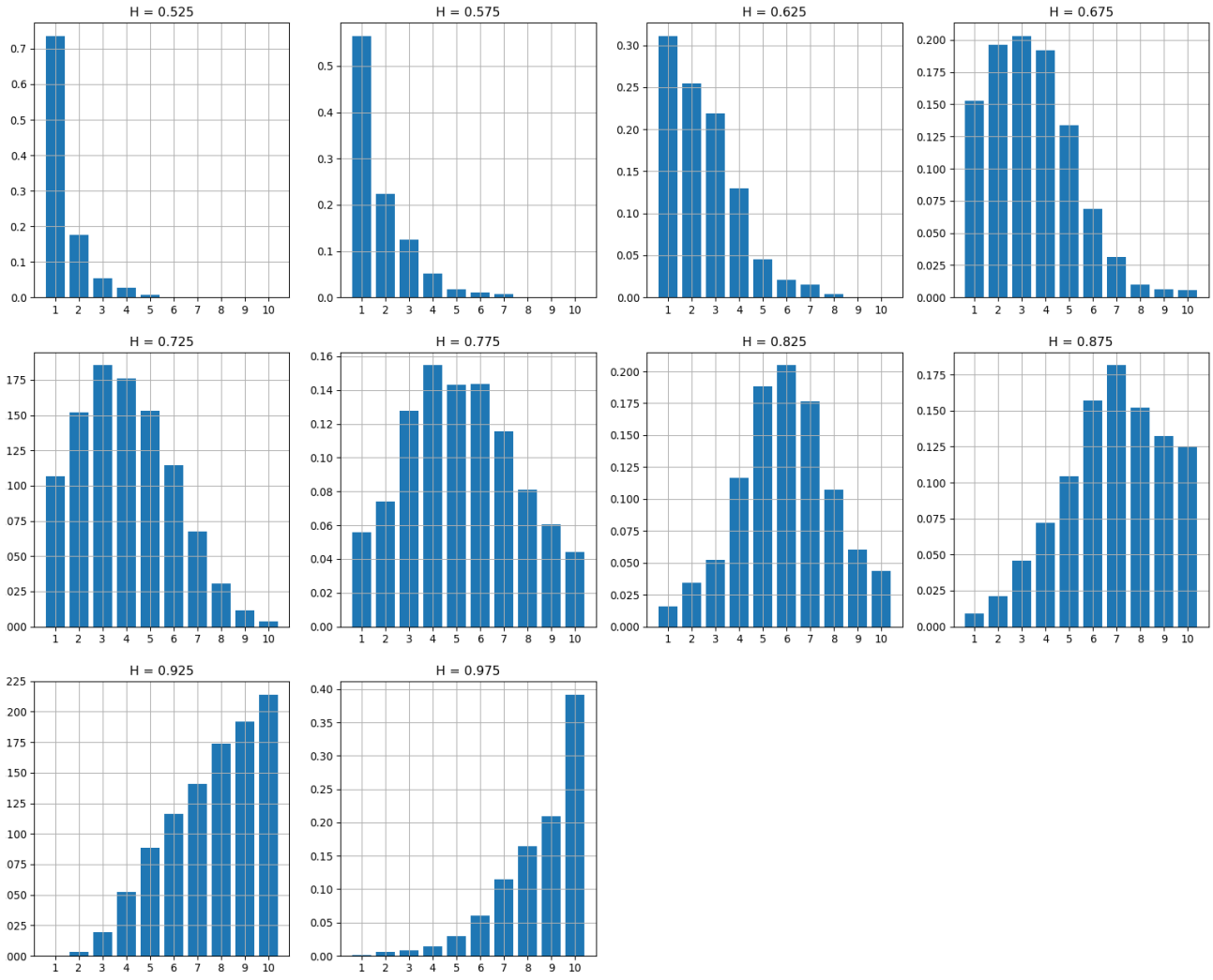

Рисунок 3.7 – Оцінювання показника Херста методом мультифрактального детрендованого флуктуаційного аналізу

Таблиця 3.24 – Порівняння ширини 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

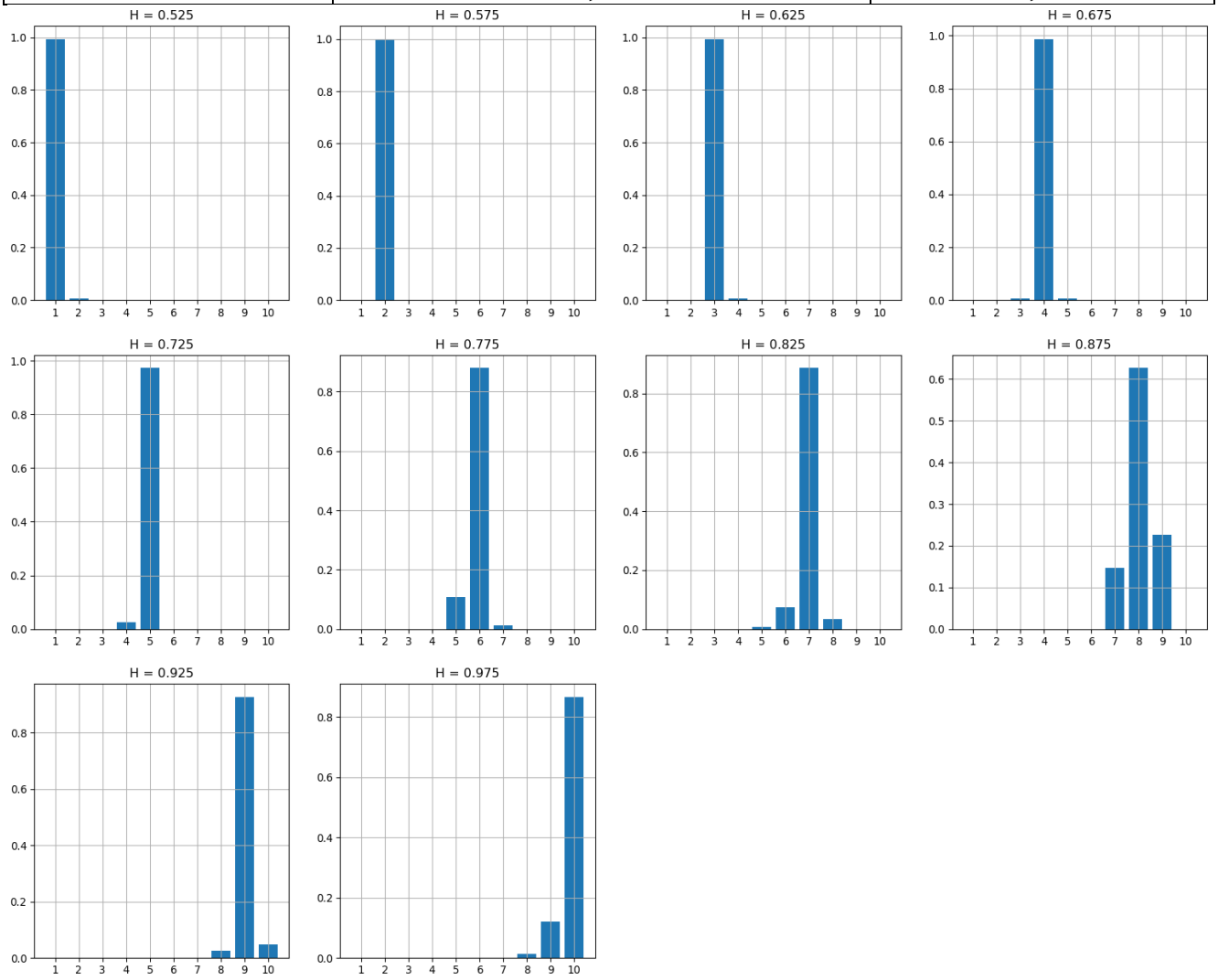

Рисунок 3.8 – Результати оцінювання показника Херста за допомогою нейронних мереж довгої короткострокової пам’яті

ВИСНОВКИ ПО ТРЕТЬОМУ РОЗДІЛУ

1. Класифікація УМД за їхніми фрактальними властивостями базується на том факті, що змінення фрактальних властивостей УМД безпосередньо пов’язано

зі зміною його кореляційної та спектральної структури та ряду статистичних властивостей. Таким чином, клас УМД з близькими фрактальними властивостями є класом об’єктів із певним комплексом спектрально-кореляційних і статистичних властивостей.

2. Здійснено чисельні експерименти класифікації різних типів упорядкованих даних, при цьому множина УМД розбивалась на класи за їхніми фрактальними властивостями. Кожен клас представляв собою набір згенерованих УМД з показником Херста, котрий лежить у вузькому діапазоні значень. Для генерації УМД в середині кожного класу значення показника Херста вибиралося за допомогою рівномірного закону розподілення.

3. Для класифікації упорядкованих масивів даних дослідним шляхом були обрані наступні методи машинного навчання: Baggins з деревами класифікації та регресії; Випадковий ліс з деревами класифікації та регресії; повнозв’язні нейронні мережі; нейронні мережі довгої короткострокової пам'яті

4. Для проведення класифікації використовувались три типи ознак для методів машинного навчання: значення упорядкованих масивів даних; статистичні та фрактальні характеристики; рекурентні характеристики. 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 монофрактальными властивостями.

5. 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.

Результати третього розділу опубліковані в [58, 60, 61, 89, 90, 117, 118,119].

РОЗДІЛ 4
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСИФІКАЦІЇ ФРАКТАЛЬНИХ УПОРЯДКОВАНИХ МАСИВІВ ДАНИХ

4.1. Розробка інформаційної технології класифікації фрактальних упорядкованих масивів даних

У роботі запропонована інформаційна технологія (ІТ) класифікації фрактальних упорядкованих масивів данних. Проведена апробація запропонованої ІТ на віртуальній мережі з використанням відкритого набору даних мережного трафіку [106, 122].

На рисунках 4.1, 4.2 представлена UML-модель запропонованої ІТ класифікації фрактальних упорядкованих масивів даних. Вона складається з 8 основних функціональних частин, а саме:

- Збір інформації
- Збереження вхідної інформації до БД, для того щоб можна було використовувати її повторно
- Обробка вхідної інформації запропонованим алгоритмом для підготовки даних котрі будуть використовуватися для навчання моделі та генерації управління на основі навченої моделі
- Навчання моделі та збереження її до бази даних
- Класифікація вхідної інформації та генерація управління
- Використання результатів роботи моделі
- Аналіз ефекту від використання результатів роботи моделі
- Представлення результатів функціонування інформаційної технології усереднених за деяких час її використання

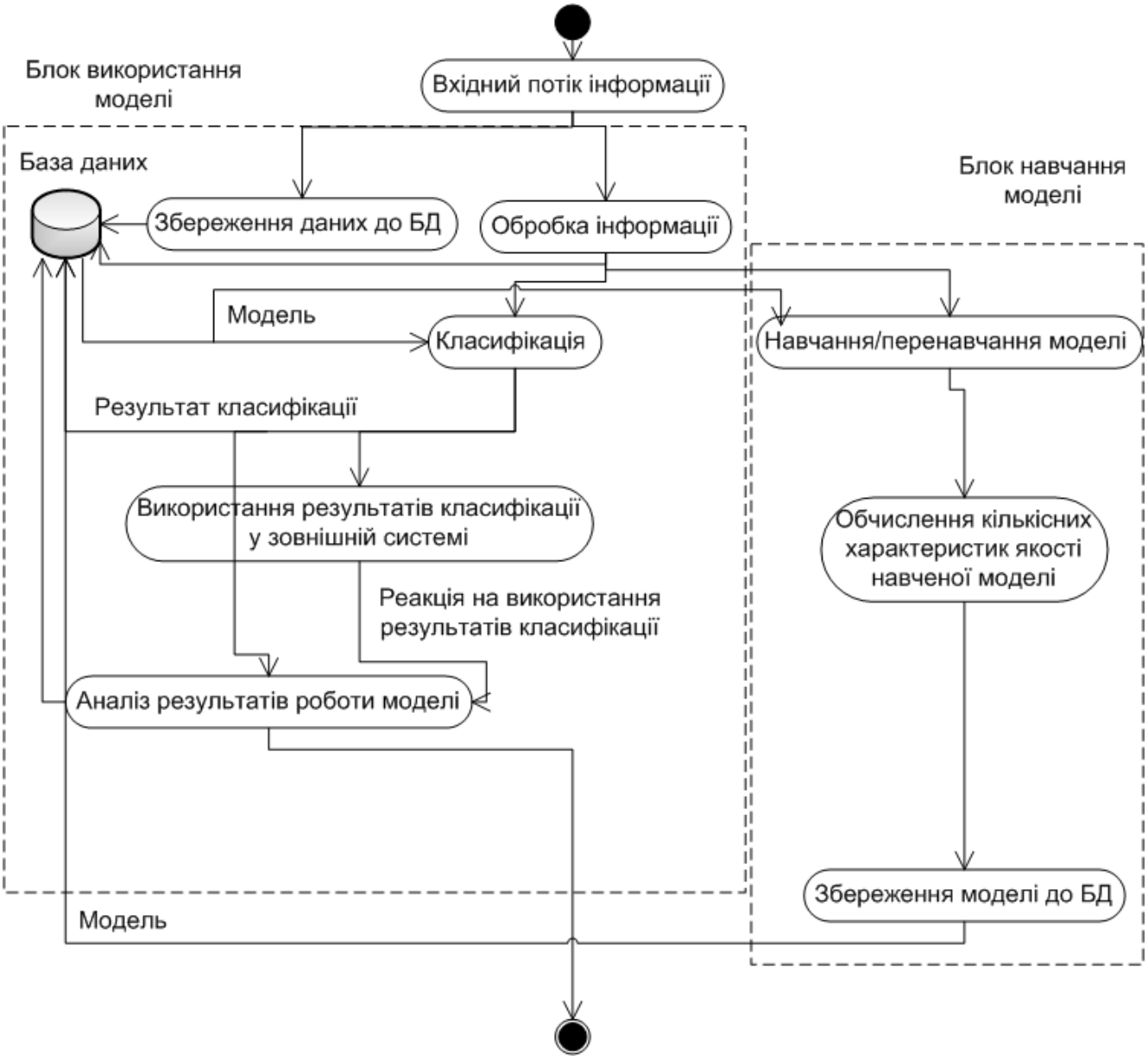

Рисунок 4.1 – Схема інформаційної технології класифікації упорядкованих масивів даних

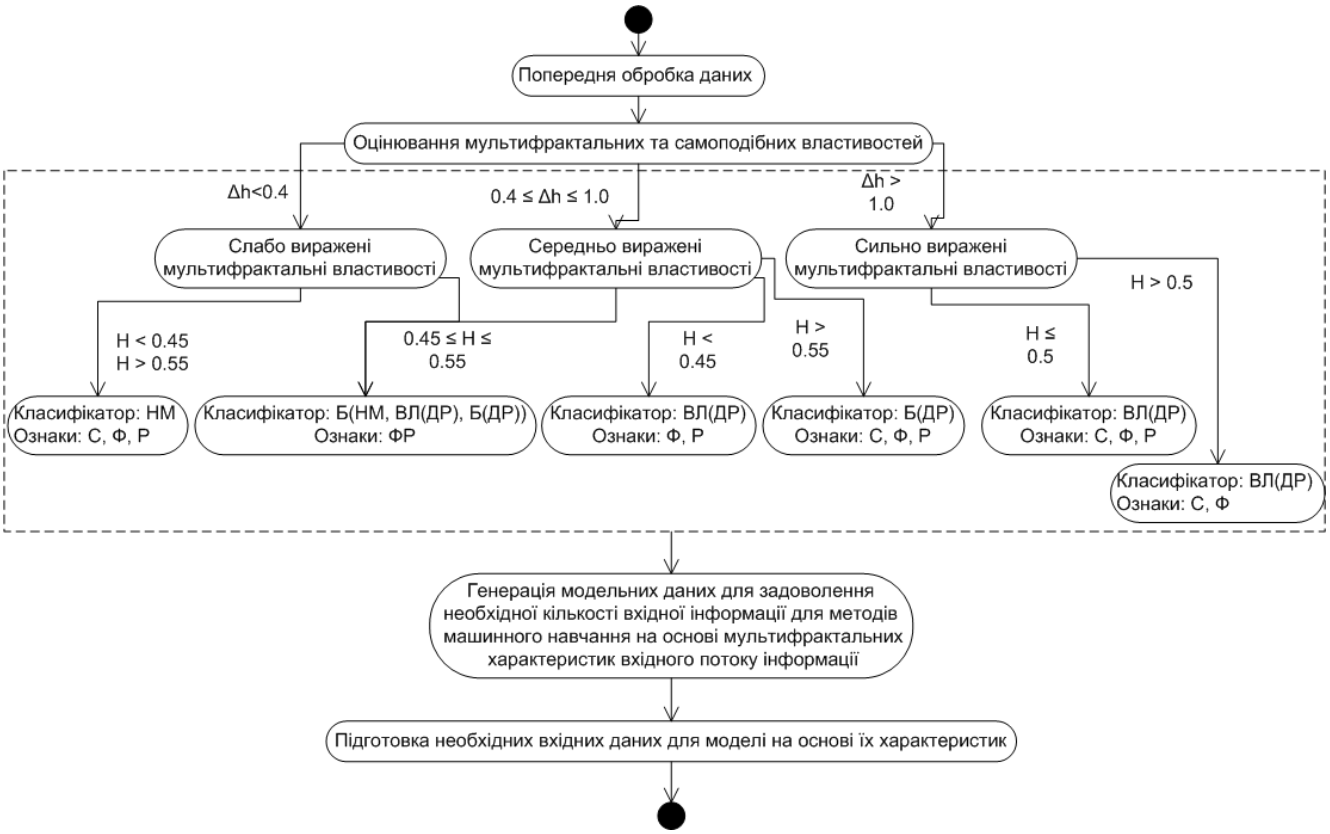

Рисунок 4.2 – Декомпозиція блоку «Обробка інформації»

Розглянемо короткий опис ІТ:

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

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

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

2.2. Оцінювання самоподібних і 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.

2.3. Аналіз самоподібних властивостей УМД дозволяє виділити три умовних діапазону для показника Херста *H*: антиперсистентність при *H*<0,45, персистентність при значеннях *H*>0,55 і фактична відсутність довгострокової залежності $0,45 \le H \le 0,55$. У даному випадку, враховуючи похибки оцінювання показника Херста, ми 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 діапазону мультифрактальності для функції $h(q)$: слабкі мультифрактальні властивості $\Delta h(q) < 0,4$, середні мультифрактальні властивості $0,4 \le \Delta h(q) \le 1$ і сильну мультифрактальність $\Delta h(q) > 1$. Дослідження показали, що діапазон мультифрактальних і самоподібних властивостей УМД відіграє важливе значення для вибору класифікатора і, відповідно, точності класифікації. Дослідження показали, що найбільш добре піддаються класифікації УМД з сильними мультифрактальними властивостями. В 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 можна використовувати самі значення УМД. Це займає набагато більш часу, але не потребує ніякого оцінювання фрактальних характеристик.

2.5. У разі, коли УМД явно має мультифрактальні властивості, але діапазон узагальненого показника Херста має деяке невелике значення $0,4 \le \Delta h(q) \le 1$, 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 можна вважати умовно монофрактальним, то як додаткові ознаки 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 ($0,45 \le H \le 0,55$). 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 інформаційної технології в умовах недостатньої кількості інформації.

3. Навчання/перенавчання моделі. Цей блок інформаційної технології 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, цей модуль необхідний для розуміння ступінь точності моделі.

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

6. Класифікація, видача керування. 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 модуль та ті що йдуть за ним 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 інформаційної технології та генерації звітів.

4.2.Застосування запропонованої ІТ на прикладі детектування DDoS-атак

Розроблена технологія класифікаццї фрактальних масивів даних була використана для детектування DDoS-атак[1, 7, 54, 69, 77]. DDoS-атака – хакерська атака на обчислювальну систему з метою створення таких умов, при яких користувачі системи не можуть отримати доступ до системних ресурсів. В даний час DDoS-атаки дозволяють довести до відмови практично будь-яку систему, не залишаючи юридично значимих доказів. Схема адаптованої інформаційної технології для виявлення DDoS-атак представлена на рисунку 4.3.

Account Takeover (ATO) – це атака великого витоку даних користувачів(наприклад, автоматичне тестування імен користувачів/паролів по списку вкрадених даних) цей тип атаки в часових рядах проявляється як переривчасті сплески активності, які розповсюджуються з часом. Data Exfiltration-це атака витікання даних(наприклад, передача даних у хмарне середовище). При налаштуванні багатомірного вектору на окремій точці по запитам – багатомірні часові ряди.

Одним із рішень задачі своєчасного виявлення атаки є розробка класифікатора, який би визначав ймовірність, що трафік який приходить є атакою. Останні дослідження показують, що однією з характерних ознак атаки є зміна показника Херста у трафіка, що містить атакуючі файли [24, 33, 35].

Для виявлення вищевказаних атак також застосовуються алгоритми виявлення аномалій для часових рядів [43, 45, 66, 70, 79]. Алгоритм виявлення аномалій повинен:

1. Маркувати кожен момент часу аномалія/ не аномалія
2. Прогнозувати сигнал для деякої точки та виявляти ступіть змінення значення точки від прогнозованого значення, щоб рахувати його аномалією або ні.

Проблема виявлення аномалій для часових рядів за звичай формується як пошук точок даних більш високого рівня відносно деякого середнього рівня сигналу. За звичай розглядають несподівані сплески, падіння, зміни тренду та зрушення рівнів [80, 85, 92, 93].

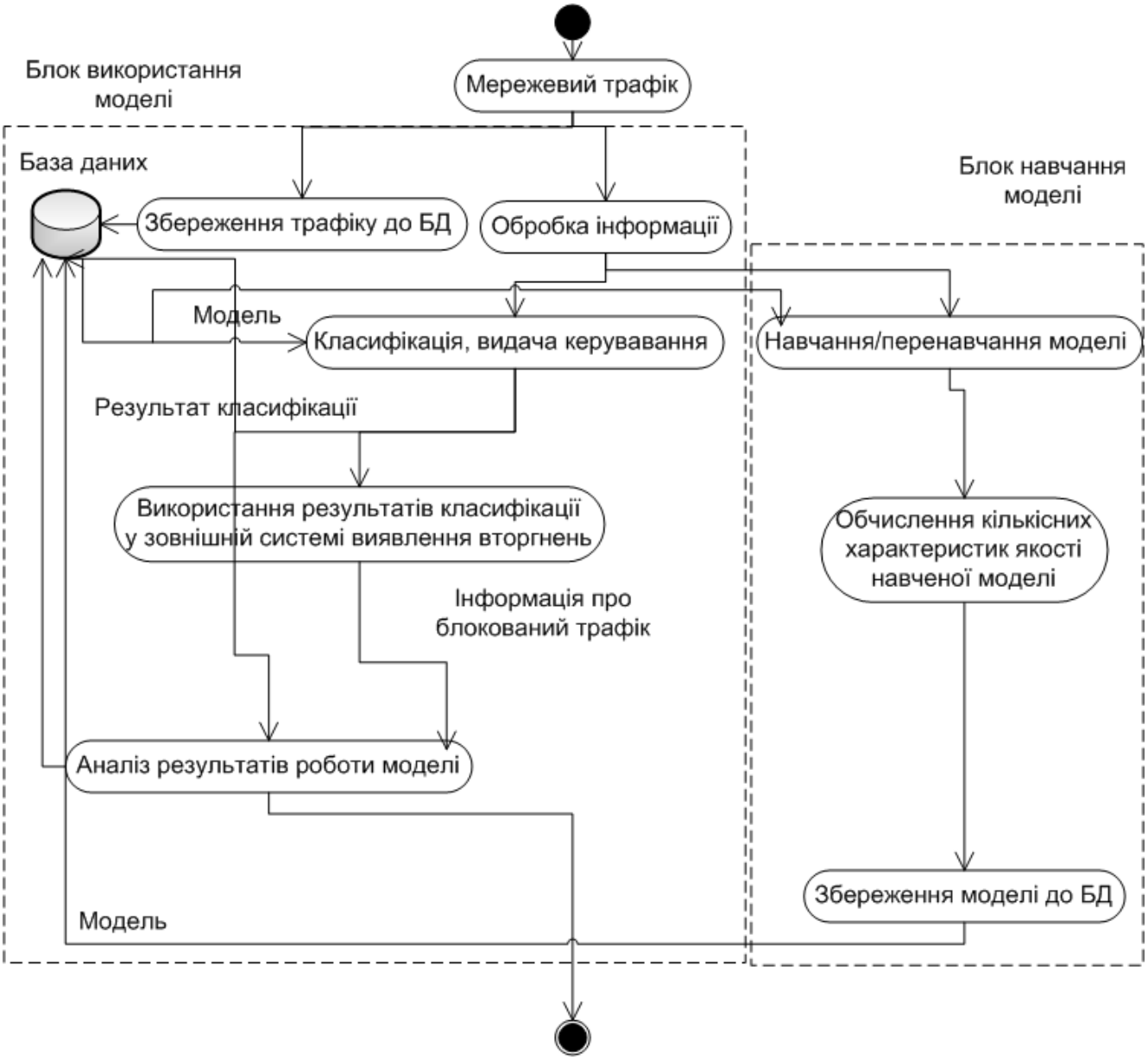

Рисунок 4.3 – Схема адаптованої інформаційної технології для виявлення DDoS-атак

Продуктивність та ефективність систем виявлення вторгнень (СВВ) Intrusion Detection System (IDS) оцінюється за допомогою різноманітних параметрів, таких як вартість, використання ресурсів, швидкість, але якщо IDS може дати

класифікацію атакам та звичайним подіям, то це являється спостерігаючим параметром [28, 32, 34, 37]. В залежності від характеру цієї події та виявлення із IDS використовується чотири можливих результати:

1. True positive (TP): процеси які фактично атакують і успішно виявляються та помічені як атаки.
2. False positive (FP): нормальний процес класифікується як атака.
3. True negative (TN): процеси які насправді являються нормальними та успішно помічені як нормальні.
4. False negative (FN): це атаки що неправильно класифікуються, як нормальні події.

False Positive Rate (FPR) вказує на пропорцію в якій законна інформація помилково виявляється як зловмисна. Зрозуміло що високий FPR визиває менш ефективну IDS, а високий FNR зробить систему вразливою для вторгнення. TNR дає інформацію про кількість виявлених атак із всіх подій атаки. У подальшому, точність по відношенню до спільних подій відносить до долі подій, що виявлені або класифіковані як точні [37, 39]. Таким чином, щоб максимізувати продуктивність IDS, rates FP та FN повинні бути мінімізовані при максимальній точності [49, 50]. Всі підходи що пропонувалися для зменшення помилкових спрацьовувань, не зовсім вірні по причині того, що просто зменшення помилкових спрацьовувань недостатнє. Тобто, необхідно зменшити FPR, при цьому зберігаючи точність на одному рівні або збільшуючи її до певного рівня.

Для проведення експериментів були взяті дані трафіків з двох наборів реальних даних. З першого набору були взяті реалізації DDoS-атак, а з другого – дані трафіка реальної мережі Internet Service Provider на канальному рівні [67, 101]. Реалізація трафіка під дією DDoS -атаки є сумою трафіка і реалізації одного з видів атаки. В ході роботи проведені експерименти по обчисленню зміни узагальненого показника Херста під дією DDoS-атак. На рис. 4.4 показано мережний трафік, реалізацію атаки та сумарний атакований трафік [40, 55].

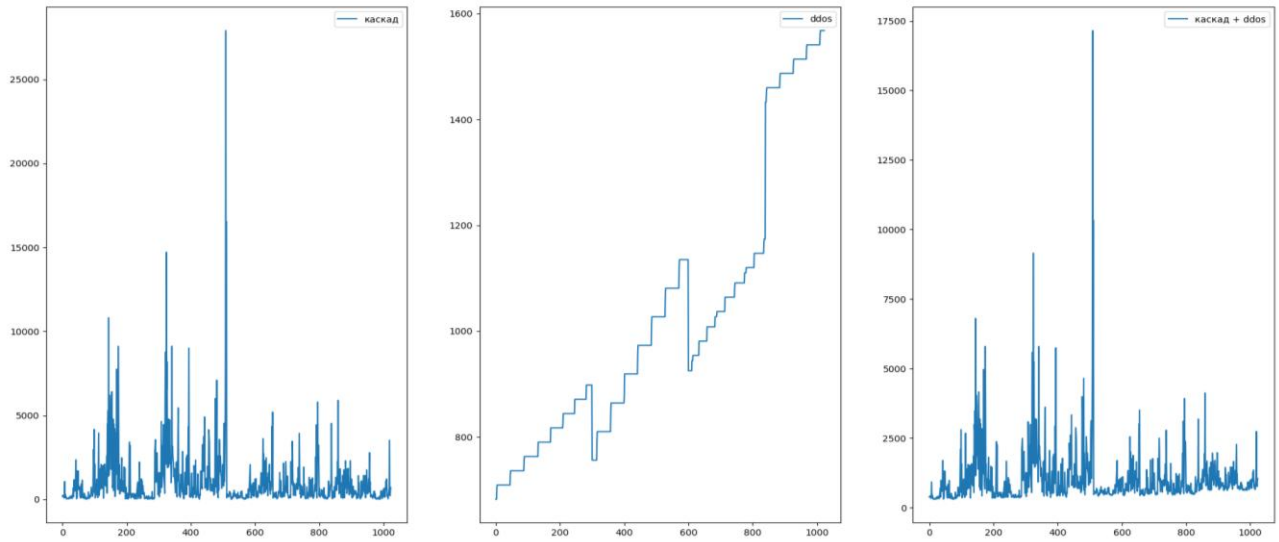

Рисунок 4.4 – Мережний трафік(ліворуч) часовий ряд атаки(посередині) та сумарний часовий ряд (справа)

У таблиці 4.1 представлені фрактальні властивості для різних нормальних і атакованих 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 і сильні мультифрактальні властивості.

Таблиця 4.1 – 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

Таким чином ми маємо провести бінарну класифікацію персистентих УМД з сильно вираженими мультифрактальними властивостями. З урахуванням запропонованої ІТ методом класифікації був обраний випадковий ліс на основі регресійних дерев рішень. В якості ознак були використані статистичні та фрактальні характеристики реалізацій трафіку. Результатом роботи моделі була ймовірність відповідності реалізації трафіку заданому класу (є атака, чи немає атаки).

На етапі навчання класифікаторів інформаційної технології на вхід ІТ подавалися атакований трафік та момент часу початку і кінця DDoS -атаки. На цьому етапі проводиться збір даних, обчислюються необхідних для методів машинного навчання характеристики з подальшим навчанням класифікатора виявлення DDoS -атак у трафіку. Навчальна вибірка створювалася методом взяття ділянки довжиною 120 секунд, з кроком 10 секунд трафіка з потоку даних, при цьому ділянка трафіка маркувалася присутністю або відсутністю на ньому DDoS - атаки. Довжина ділянки трафіка з DDoS атакою була вибрана рівною $\frac{1}{4}$ довжини всієї реалізації. Положення ділянки трафіка з атакою вибиралося випадково. На рисунку 4.5 показано приклад, де довжина трафіка була 4096 значень, довжина ділянки трафіка з атакою – 1024 інтервал ділянки атакою – 2433-3457.

Для класифікації атакованого та не атакованого трафіку були використані статистичні та мультифрактальні характеристики, отримані по значенням часового ряду: середньоквадратичне відхилення, максимальне значення, та медіана реалізації трафіка, середнє значення, середньоквадратичне відхилення та діапазон узагальненого показника Херста, значення показника Херста H.

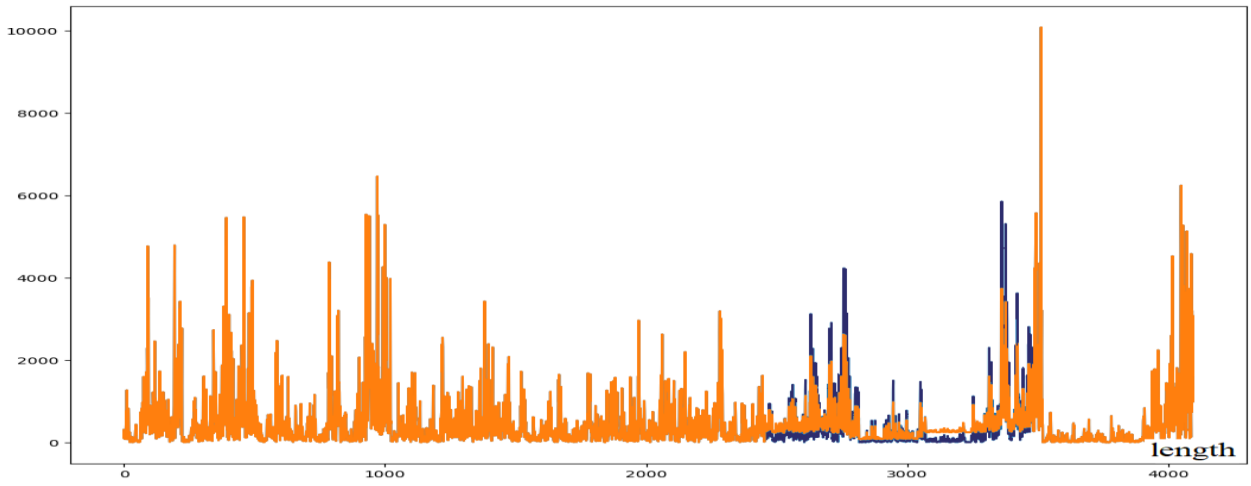

Рисунок 4.5 – Атакований трафік, інтервал ділянки з атакою 2433-3457

При наборі достатньої кількості прикладів різних класів ІТ переходить до етапу генерації навчальної вибірки з доповненням даних [108, 124, 128], методами генерації, до достатньої кількості навчальних прикладів для методів машинного навчання, після чого можна приступити до вибору характеристик трафіка які будуть використовуватися та навчання необхідної моделі класифікації. Після завершення навчання, обчислюються статистичні характеристики моделі на навчальній вибірці, після чого модель зберігається до бази даних.

Після того як модель класифікації була навчена її можна використовувати для виявлення DDoS -атак у вхідному потоку даних. Для цього необхідно перестати відправляти моменти часу початку та кінця DDoS-атаки в ІТ, а передавати лише мережний трафік. Після детектування DDoS-атаки передається керуючий сигнал зовнішній системі [46, 51], в нашому випадку це модуль який передає трафік у ІТ, після чого зовнішня система інформує інформаційну технологію про наявність або відсутність атаки в цей момент часу. Заключним етапом роботи інформаційної технології є збір статистичних оцінок результатів використання ІТ, де показується кількість виявлених атак, а також кількість хибних тривог, з можливістю вибрати інтервал часу відображення статистики, відображення на часовій вісі або в якості табличних значень.

Для оцінки якості роботи інформаційної технології була розгорнута віртуальна мережа [44, 97], в якій трафік передавався з п’яти джерел на сервер

ідентифікації атаки. На окремому сервері була розгорнута запропонована інформаційна технологія на якому проводилося навчання моделей класифікації з подальшим їх використанням для визначення атак в мережі. Весь 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 таблиці 4.2 представлено показники якості роботи запропонованої інформаційної технології 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 (FN+TN), F-міра.

При порівнянні якості роботи запропонованої ІТ виявлення вторгнень з існуючими методами визначено [16, 38, 78, 95, 110], що запропонований метод краще на 1-2% ідентифікує атаки, при цьому відсоток невиявлених атак менший на 1-2%, помилкових спрацьовувань більше на 0.2 – 0.5% та час ідентифікації атак відносно однаковий при роботі існуючих методів.

Таблиця 4.2 – 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
F1 міра	0,977	0,956

ВИСНОВКИ ПО ЧЕТВЕРТОМУ РОЗДІЛУ

1. 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.

2. Запропонована технологія була застосована для детектування DDoS-атак на обчислювальну систему. Був розроблений класифікатор, який визначає ймовірність того, що аналізований трафік є атакованим. Для проведення експериментів були застосовані реалізації DDoS-атак, трафіків реальної мережі та модельні реалізації трафіків з різними фрактальними властивостями.

3 Для детектування DDoS-атак були запропоновано методи бінарної класифікації часових рядів з фрактальними властивостями, що засновані на використанні дерев прийняття рішень та нейронних мереж, які використовують як ознаки фрактальні та рекурентні характеристики, що дозволило збільшити точність класифікації.

4. Для оцінки якості роботи запропонованої інформаційної технології була розгорнута віртуальна мережа. Приведені оцінки якості роботи мережі з впровадженою технологією детектування атак, які свідчать про її переваги.

Результати четвертого розділу опубліковані в [57, 87, 88, 91, 106, 121, 122, 131].

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 монофрактальными 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 кібер-атак 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 ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Le A., Cheriton D.R., Boutaba R., Al-Shaer E. Correlation-based load balancing for network intrusion detection and prevention systems. *4th international ICST conference on security and privacy in communication networks*, september 2008. doi: https//doi.org/10.1145/1460877.1460880

2. Gibson A., Patterson J. Deep Learning. O'Reilly Media, Inc., 2017. URL: https://www.oreilly.com/library/view/deep-learning/9781491924570/

3. Abry P., Goncalves P., Vehel Scaling J. L. Fractals and Wavelets London. John Wiley & Sons. 2009. 464 p.

4. Abry P., Veitch D. Wavelet analysis of long-range dependent traffic. *IEEE/ACM Transactions Information Theory*. 1998. № 1 (44). P. 2–15.

5. Abry P., Flandrin P., Taqqu M.S., Veitch D. Self-similarity and long-range dependence through the wavelet lens. *Theory and applications of long-range dependence*. Birkhäuser. 2003.

6. Ahn K.I., Lee K. Identification of nonstandard multifractional brownian motions under white noise by multiscale local variations of its sample paths. *Mathematical Problems in Engineering*. 2013. doi: https://doi.org/10.1155/2013/794130

7. Al-kasassbeh M., Al-Naymat G., Al-Hawari E. Towards Generating Realistic SNMP-MIB Dataset for Network Anomaly Detection. *Int. J. Comput. Sci. Inf. Secur*. 2016. Vol. 14. P. 1162–1185.

8. Andre L., Coelho V., Clodoaldo A., Lima M. Assessing fractal dimension methods as feature extractors for EMG signal classification. *Engineering Applications of Artificial Intelligence*. 2014. Vol. 36. P. 81–98.

9. Arjunan S. P., Kumar D. K., Naik G. R. A machine learning based method for classification of fractal features of forearm sEMG using Twin Support Vector Machines. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*. 2010. doi: https://doi.org/10.1109/IEMBS.2010.5627902

10. Ayache A., Cohen S., Véhel J.L. The covariance structure of multifractional Brownian motion, with application to long range dependence. *In: 2000*

IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, 2000. Vol. 6. doi: https://doi.org/10.1109/ICASSP.2000.860233

11. Bacry E., Delour J., Muzy J. F. A multifractal random walk. *Phys. Rev. E.* 2003. Vol. 64. Issue 2. P. 103–106.

12. Bacry E., Delour J., Muzy J. F. A multivariate multifractal model for return fluctuations. *Submitted to Phys. Rev. Lett*. 2000. URL: http://arxiv.org/pdf/cond-mat/0009260.pdf, last accessed 03.03.2021.

13. Bagnall A., Bostrom A., Large J., Lines J. Simulated Data Experiments for Time Series Classification Part 1: Accuracy Comparison with Default Settings. URL: https://arxiv.org/abs/1703.09480v1, last accessed 03.03.2021.

14. Ben D. Feature-Based Time-Series Analysis. 2017. URL: https://arxiv.org/abs/1709.08055, last accessed 03.03.2021.

15. Bianchi S. Pathwise identification of the memory function of multifractional Brownian motion with application to finance. *International Journal of Theoretical and Applied Finance*, 2005. Vol. 8(02), P. 255-281. doi: https://doi.org/10.1142/S0219024905002937

16. Binghao Y., Guodong H. LA-GRU: Building Combined Intrusion Detection Model Based on Imbalanced Learning and Gated Recurrent Unit Neural Network. *Security and Communication Networks*, 2018. Vol. 2018. 13 p. doi: https://doi.org/10.1155/2018/6026878.

17. Brambila F. Fractal Analysis – Applications in Physics. Engineering and Technology.Mexico, 2017. URL: https://www.intechopen.com/books/fractal-analysis-applications-in-physics-engineering-and-technology, last accessed 03.03.2021.

18. Breiman L. Bagging predictors. *Machine Learning*. 1996. Vol. 24, № 2. P. 123–140. URL: http://citeseer.ist.psu.edu/breiman96bagging.html, last accessed 03.03.2021.

19. Breiman L. Random Forests. *Machine Learning*. 2001. Vol. 45 (1). P. 5–32.

20. Bulakh V., Kirichenko L., Radivilova T. Classification of Multifractal Time Series by Decision Tree Methods. *Proceedings of the 14th International*

Conference ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. May 14–17 2018. Kyiv, Ukraine. 2018. Vol. I(2105). P. 457–460.

21. 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.

22. 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.

23. Buza K. Time series classification and its applications. *In: 8th International Conference on Web Intelligence, Mining and Semantics Proceedings*, 2018. P.1-4. doi: https://doi.org/10.1145/3227609.3227690

24. Kayataş C.E., Fouladi R.F., Ermiş O., Anarim E. Statistical measures: Promising features for time series based DDoS attack detection. 2018 *26th Signal Processing and Communications Applications Conference (SIU)*. Izmir. 2018. P. 1-4. doi: https://doi.org/10.1109/SIU.2018.8404348

25. Calvet L., Fisher A., Mandelbrot B. Large deviation and the distribution of price changes. *Cowles Foundation Discussion Paper*. Yale University. 1997. Vol. 1165. P. 1–28.

26. Ching E.S., Tsang Y.K. Multifractality and scale invariance in human heartbeat dynamics. *Physical Review*, 2007. Vol. 76. doi: https://doi.org/10.1103/PhysRevE.76.041910

27. Cielen D., Meysman A., Ali M. Introducing data science: big data, machine learning, and more, using Python tools. *Manning Publications Co.* 2016. 320 p.

28. Ciza T. Improving intrusion detection for imbalanced network traffic. *Security and communication Networks.* 2013. Vol. 6. P. 309–324. doi: https://doi.org/10.1002/sec.564

29. Clegg R.G. A practical guide to measuring the hurst parameter. *International Journal of Simulation. Systems, Science & Technology*. 2006. Vol. 7, No 2. P. 3–14.

30. Cont R., Tankov P. Financial modelling with jump processes. Chapman & Hall CRC Press. 2004. 552 p.

31. Corlay, S., Lebovits, J., Véhel, J. L. Multifractional stochastic volatility models. *Mathematical Finance*. 2014. Vol. 24(2). P. 364-402. doi: https://doi.org/10.1111/mafi.12024

32. Czarkowski M., Kaczmarek S., Wolff M. Influence of Self-Similar Traffic Type on Performance of QoS Routing Algorithms. *INTL Journal of electronics and telecommunications*. 2016. Vol. 62, No 1. P. 81–87.

33. Ageyev D., Qasim N. LTE EPS network with self-similar traffic modeling for performance analysis. *Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*. Kharkiv, 2015. P. 275-277.

34. Cieslak D.A., Chawla N.V., Striegel A. Combating imbalance in network intrusion datasets. *IEEE International Conference on Granular Computing*, GrC Atlanta, Georgia, USA, May 10-12, 2006. doi: https://doi.org/10.1109/GRC.2006.1635905

35. Ageyev D.V., Salah M.T. Parametric synthesis of overlay networks with self-similar traffic. *Telecommunications and Radio Engineering*. 2016. Vol. 75(14). P. 1231-1241.

36. Ali M., Way J., Scargle D., Srivastava N. Advances in Machine Learning and Data Mining for Astronomy. Manning Publications. Chapman and Hall/CRC. 2016. 744 p.

37. Deka R, Bhattacharyya D. Self-similarity based DDoS attack detection using Hurst parameter. *Secur Commun Netw*. 2016. Vol. 9(17). P. 4468–4481. dois: https://doi.org/10.1002/sec.1639

38. Dominique T., Shipmon M., Jason M., Gurevitch P. Piselli and Steve Edwards. Time Series Anomaly Detection. Detection of Anomalous Drops with Limited

Features and Sparse Examples in Noisy Highly Periodic Data, Google, Inc. Cambridge, MA, USA, 2016. P. 1-9.

39. Droniuk I., Fedevych O. Computer network protocol analyzer designed for accuracy of traffic trends forecasting. *Матеріали XII Міжнародної науково-технічної конференції Перспективні технології і методи проектування МЕМС (MEMSTECH 2016)* Львів-Поляна. 2016. С. 155–157.

40. Esling P., Agon C. Time series data mining. *ACM Computing Surveys*. 2012. Vol. 46, No. 1. P. 12-46.

41. Fauth A., Tudor C. A. Multifractal random walks with fractional Brownian motion via Malliavin calculus. *IEEE Transactions on Information Theory*. 2014. Vol. 60(3). P. 1963-1975.

42. Fawaz H.I., Forestier G., Weber, J., Idoumghar L., Muller P.A. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*. 2019. Vol. 33(4), P. 917-963.

43. Fedevych O., Droniuk I., Nazarkevych M. Monitoring and analysis of measured and modeled traffic of TCP/IP Networks. *Communications in Computer and Information Science, Springer International Publishing Switzerland*. 2016. Vol. 608. P. 32-41.

44. Jiang H., Zhang G., Xie G., Salamatian K., Mathy L. Scalable high-performance parallel design for Network Intrusion Detection Systems on many-core processors. *Architectures for Networking and Communications Systems*, San Jose, CA, 2013. P. 137-146. doi: https://doi.org/10.1109/ANCS.2013.6665196

45. Suda H., Natsui M., Hanyu T. Systematic Intrusion Detection Technique for an In-vehicle Network Based on Time-Series Feature Extraction. *IEEE 48th International Symposium on Multiple-Valued Logic (ISMVL)*, Linz, 2018. P. 56-61. doi: https://dois.org/10.1109/ISMVL.2018.00018

46. Haproxy T. ALOHA load balancer Stateful firewalls, IPS, IDS and UTM load balancing. 2019. URL: ttps://www.haproxy.com/support/technical-notes/an-0062-en-stateful-firewalls-ipsids-and-utm-load-balancing/, last accessed 03.03.2021.

47. Harikrishnan K.P., Misra R., Ambika G. Can the multifractal spectrum be used as a diagnostic tool ? *Chaotic Modeling and Simulation*. 2013. Vol. 1. P. 51–57.

48. Ioffe S. Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *In: 32nd International Conference on Machine Learning Proceeding*, PMLR 37, Lille, France, 2015. P. 448-456. URL: https://arxiv.org/abs/1502.03167, last accessed 03.03.2021.

49. Jabeza J., Muthukumar B. Dr. Intrusion Detection System (IDS): Anomaly Detection Using Outlier Detection Approach. *Procedia Computer Science*. 2015. Vol. 48. P. 338-346. doi: https://doi.org/10.1016/j.procs.2015.04.191

50. Lee J., Park K. GAN-based imbalanced data intrusion detection system. *Personal and Ubiquitous Computing*. 2019. P. 1-8. doi: https://doi.org/10.1007/s00779-019-01332-y

51. Kang J., Yang M., Zhang J. Accurately Identifying New QoS Violation Driven by High-Distributed Low-Rate Denial of Service Attacks Based on Multiple Observed Features. *Journal of Sensors*. 2015. Vol. 11. doi: https://doi.org/10.1155/2015/465402.

52. Kantelhardt J.W., Koscielny-Bunde E., Havlin S. Detecting long-range correlations with detrended fluctuation analysis. *Physica A. Statistical Mechanics and its Applications.* 2001. Vol. 295, No 3–4. P. 441–454.

53. Kantelhardt J.W. Fractal and Multifractal Time Series. 2008. URL: http://arxiv.org/abs/0804.0747, last accessed 03.03.2021.

54. Kaur G, Saxena V, Gupta J. Detection of TCP targeted high bandwidth attacks using self-similarity. J King Saud Univ Comput Inf Sci. 2017. doi: https://doi.org/10.1016/j.jksuci.2017.05.004

55. Khor K.C., Ting C.Y., Phon-Amnuaisuk S. The Effectiveness of Sampling Methods for the Imbalanced Network Intrusion Detection Data Set. In: Herawan T., Ghazali R., Deris M. (eds) *Recent Advances on Soft Computing and Data Mining*. Advances in Intelligent Systems and Computing, Springer, Cham. 2014. Vol. 287. P. 613-622. doi: https://doi.org/10.1007/978-3-319-07692-8_58

56. Kingma D.P., Ba J. Adam. A Method for Stochastic Optimization. *In: 3rd International Conference on Learning Representations (ICLR) Proceeding*, San Diego, USA, 2015. URL: https://arxiv.org/abs/1412.6980, last accessed 03.03.2021.

57. 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 міжнар. наук.-практ. конф. 8-10 листоп. 2018 р. Чернівці, 2018. С. 13.

58. 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.

59. 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.

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

61. 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.

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

63. Korus L., Piorek M. Compound method of time series classification. *Nonlinear Anal Model Control.* 2015. Vol. 20(4). P. 545–560. doi: https://doi.org/10.15388/NA.2015.4.6

64. Buza K. Time Series Classification and its Applications WIMS '18. *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics.* 2018. doi: https://doi.org/10.1145/3227609.3227690

65. Ruiz S. E., García G., Aviña G., Hernández D. Analysis of self-similar data by artificial neural networks. *In: Proceedings of the 2011 International Conference on Networking, Sensing and Control*, Delft, 2011. P. 480-485. doi: https://doi.org/10.1109/ICNSC.2011.5874873

66. Leland W.E., Taqqu M.S., Willinger W., Wilson D.V. Statistical analysis of high time-resolution Ethernet LAN traffic measurements. *Computing Science and Statistics*. 1993. Vol. 25. P. 146–155.

67. Li M., Zhao W., Chen S. MBm-based scalings of traffic propagated in internet. *Mathematical Problems in Engineering.* 2011. doi:10.1155/2011/389803

68. Zhang Y., Xiong R., Hongwen H., Michael G., Pecht Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. 2018. URL: https://ieeexplore.ieee.org/abstract/ document/8289406, last accessed 03.03.2021.

69. Zhou L., Liao M., Yuan C., Zhang H. Low- Rate DDoS Attack Detection Using Expectation of Packet Size. *Security and Communication Networks*. Vol. 2017. 14 p. doi: https://doi.org/10.1155/2017/3691629.

70. Andreolini M., Casolari S., Colajanni M., Marchetti M. "Dynamic load balancing for network intrusion detection systems based on distributed architectures". *Sixth IEEE International Symposium on Network Computing and Applications* (NCA 2007), Cambridge, MA, 2007, P. 153-160. doi: http://doi.org/10.1109/NCA.2007.17

71. Makowiec D., Fulinski A. Multifractal detrended fluctuation analysis as the estimator of long-range dependence. *Acta Physica Polonica B*. 2010. Vol. 41. № 5. P. 1025-1050.

72. Mandelbrot B, Van Ness J.W. Fractional Brownian motions, fractional noises and applications. *SIAM Review*. 1968. Vol. 10. № 4. P. 422–437.

73. Mandelbrot B. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Mathematical Sciences Department*, IBM Thomas J. Watson Research Center, Yorktown Heights, 1974.

74. Mandelbrot B., Freeman W.H. The Fractal Geometry of Nature. San Francisco, 1983. 386 p.

75. Marwan N., Romano M., Thiel M., Kurths J. Recurrence plots for the analysis of complex system. *Physics Reports.* (2007). No 438(5–6), P. 237-329. doi: https://doi.org/10.1016/j.physrep.2006.11.001

76. Meyer Y., Sellan F., Taqqu M.S. Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion. *The Journal of Fourier Analysis and Applications*. 1995. Vol. 5, No 5. P. 465–494.

77. Idhammad M., Afdel K., Belouch M. Detection System of HTTP DDoS Attacks in a Cloud Environment Based on Information Theoretic Entropy and Random Forest. *Security and Communication Networks*. Vol. 2018, 2018. 13 p. doi: https://doi.org/10.1155/2018/1263123.

78. Gupta N., Srivastava K., Sharma A. "Reducing False Positive in Intrusion Detection System: A Survey". *International Journal of Computer Science and Information Technologies.* 2016. Vol. 7 (3). P. 1600-1603.

79. Nycz M., Nycz T., Czachorski T. Modelling dynamics of TCP flows in very large network topologies. Information Sciences and Systems 2015, *Springer International Publishing Switzerland.* 2016. Vol. 363, Part V. P. 251-259.

80. Park K., Willinger W. Self-similar network traffic and performance evaluation. *IEEE/ACM Transactions on Networking*. 2000. No 6 (3). P. 65-84.

81. Pascanu R., Gulcehre C., Cho K., Bengio Y. How to Construct Deep Recurrent Neural Networks. URL: https://arxiv.org/pdf/1312.6026.pdf , last accessed 03.03.2021.

82. Peters E. E. Fractal Market Analysis: applying chaos theory to investment and economics. J. Wiley & Sons, 1994. 336 p.

83. Prakasa R. Self-Similar Processes, Fractional Brownian Motion and Statistical Inference. 2003. URL: http://www.isid.ac.in/estatmath/eprints, last accessed 03.03.2021.

84. Pramanik S., Datta R., Chatterjee P. Self-similarity of data traffic in a Delay Tolerant Network. *In: 2017 Wireless Days*. 2017. P. 39–42. doi: https://doi.org/10.1109/wd.2017.7918112

85. Hu Q., Yu S.-Y., Asghar M. R. Analysing performance issues of open-source intrusion detection systems in high-speed networks. *Journal of Information Security and Applications*. 2020. Vol. 51.doi: https://doi.org/10.1016/j.jisa.2019.102426

86. 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.

87. 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 міжнар. наук.-практ. конф. Чернівці, 2018. С. 97.

88. 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.

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

90. 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.

91. 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. Kijv, Ukraine, 2018. P. 143–146.

92. Rao U. H., Nayak U. Intrusion Detection and Prevention Systems. *In: The InfoSec Handbook.* Apress, Berkeley, CA, 2014. P. 225-243. doi: https://doi.org/10.1007/978-1-4302-6383-8_11

93. Riedi R., Willinger W. Toward an improved understanding of network traffic dynamics. Self-Similar network traffic and performance evaluation. 2000. 574 p.

94. Rummelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation. Vol. 1 of Computational models of cognition and perception, chap. 8. Cambridge, MA: MIT Press, 1986. P. 319–362.

95. Popa S.M., Manea G.M. Using Traffic Self-Similarity for Network Anomalies Detection. *20th International Conference on Control Systems and Computer Science*, Bucharest, 2015, P. 639-644. doi: 10.1109/CSCS.2015.89

96. Shelukhin O.I. Smolskiy S.M. Osin A.V. Self-Similar Processes in Telecommunications. John Wiley & Sons, New York, USA. 2007. 320 p.

97. Rodda S., Shankar U., Erothi R. Class imbalance problem in the Network Intrusion Detection Systems. *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT).* 2016. doi: https//doi.org/10.1109/ICEEOT.2016.7755181

98. Taqqu M., Willinger M., Sherman R. Proof of a Fundamendal Result in Self-Similar Traffic Modeling. *IEEE/ACM Transactions on Networking*. 1997. Vol. 5(2). 254 p.

99. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning Data Mining, Inference, and Prediction. 2009. 745 p.

100. Wei Z., Junhao W., Min G., Haijun R., Peng L. Abnormal Profiles Detection Based on Time Series and Target Item Analysis for Recommender Systems. *Mathematical Problems in Engineering.* 2015. Vol. 2015. P. 9. doi: https://doi.org/10.1155/2015/490261.

101. Альомар М. Прогнозування самоподібного трафіка у пакетних мережах. *Управління розвитком складних систем*. Київ: КНУБА, 2014. Вип. 20. С. 102–109.

102. Андреас М. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Москва: Альфа-книга, 2017. С. 487.

103. Бессараб В.І., Ігнатенко Е.Г., Червінський В.В. Генератор самоподібного трафіку для моделей інформаційних мереж. *Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація*. Донецьк, 2008. Вип. 15 (130). С. 23-29.

104. Божокин С.В., Паршин Д.А. Фракталы и мультифракталы. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. С. 128.

105. Будкова Л.В., Корнієнко В.І. Моделювання самоподібного трафіка в інформаційних телекомунікаційних мережах. *Вісник Кременчуцького національного університету імені Михайла Остроградського*, 2013. Вип. 4. С. 46-52.

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

107. 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.

108. Дронюк І.М., Федевич О.Ю. Моніторинг та аналіз реального та модельного трафіку комп'ютерної мережі. *В зб. наук. пр. Інституту проблем моделювання в енергетиці ім.Г. Є. Пухова.* Київ, 2015. Вип.74. С. 74-81.

109. Дейнеко Ж.В., Замула А.А., Кириченко Л.О., Радивилова Т.А. Об одном методе моделирования самоподобного стохастического процесса. *Вісник Харківського національного університету ім. В. Н. Каразіна. Сер. Математичне моделювання. Інформаційні технології. Автоматизовані системи управління.* 2010. № 890. Вип. 13. С. 53–63.

110. Ильницкий С.В. Работа сетевого сервера при самоподобной (self-similar) нагрузке. Режим доступу: http://314159.ru/ilnickis/ilnickis1.pdf

111. Кириченко Л.О., Демерчян К.А., Кайали Э., Хабачёва А.Ю. Моделирование телекоммуникационного трафика с использованием стохастических мультифрактальных каскадных процессов. *Радіоелектроніка, інформатика, управління*. 2012. №1 (26). С.48-53.

112. Кириченко Л.О., Радивилова Т.А. Основные понятия, характеристики и модели самоподобных и мультифрактальных стохастических процессов. *International Journal Information content and processing*. Болгарія: Софія, 2018. Vol. 5 (1). P. 6–40.

113. Кириченко Л.О., Радивилова Т.А. Оценивание мультифрактальных характеристик стохастических процессов. *International Journal Information content and processing*. Болгарія: Софія, 2018. Vol. 5 (2). P. 106–141.

114. Кириченко Л.О., Радивилова Т.А. Оценивание параметра самоподобия для стационарных стохастических процессов. *International Journal Information content and processing*. Болгарія: Софія, 2018. Vol. 5 (1). P. 41–71.

115. Кириченко Л. О., Радивилова Т. А. Фрактальный анализ реальных данных. *International Journal Information content and processing*. Болгарія: Софія, 2018. Vol. 5 (2). P. 142–199.

116. 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.

117. Кириченко Л.О., Булах В.А. Классификация временных рядов на основе фрактальных свойств. *Міжнародний науковий симпозіум Інтелектуальні рішення*: матеріали V-ої міжнар. наук.-практ. конф. 15–20 квітня 2019 р. Ужгород, 2019. С. 220.

118. 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.

119. Кириченко Л.О., Булах В.А., Радивилова Т.А. Сравнительный анализ классификации мультифрактальных временных рядов. *Інформаційні технології в металургії та машинобудуванні*: матеріали міжнар. наук.-техн. конф. 27 – 29 березня 2018 р. Дніпро, 2018. С.126.

120. Кіріченко Л.О., Радівілова Т.А. Фрактальний аналіз самоподібних і мультифрактальних часових рядів : монографія. Харків: ФОП Панов А. Н., 2019. С. 106. URL: https://search.crossref.org/?q=10.30837%2F978-617-77-22-82-2, в останнє доступно 03.03.2021.

121. 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, last accessed 03.03.2021.

122. 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.

123. 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 міжнар. наук.-практ. конф. (1 міжнар. симпозіум). Чернівці, 2017. С. 19–20.

124. Кордяк В.І, Дронюк І.М., Федевич О.Ю. Інформаційна технологія моніторингу та аналізу трафіку у комп’ютерних мережах. *Вісник Нац. ун-ту Львівська політехніка*. 2015. № 826 : Комп'ютерні науки та інформаційні технології. С. 35–42.

125. Костромицкий А.И., Волотка В.С. Подходы к моделированию самоподобного трафика. *Вост.-Европ. журн. передовых технологий*. 2010. № 4/7. С. 46-49.

126. Кроновер Р. Фракталы и хаос в динамических системах. Москва, Постмаркет, 2000. 264 с.

127. Кучук Г.А. Аналіз та моделі самоподібного трафіка. *Авиационно-космическая техника и технология*. 2006. № 9. С. 173-180.

128. Муранов О.С., Можаев О.О., Воробйов О.В. Експериментальні дослідження механізмів прогнозування пульсацій пакетного трафіку. *Защита информации : сб. науч. трудов Национального авиационного университета*. Київ : Изд-во НАУ, 2008. Спец. вып. С. 137-142.

129. Плас Д.В. Python для сложных задач. Наука о данных и машинное обучение. Руководство. Москва: Питер, 2018. С. 759.

130. Привалов А.Ю., Баева М.В. Моделирование самоподобного трафика. *Известия Самарского научного центра Российской академии наук*, 2006. Т. 8, № 4. С. 1041-1046.

131. 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.

132. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации: Учебное пособие. М.: ФИЗМАЛИТ. 2005. 365 с.

133. Федер Е. Фракталы : пер. с англ. Москва: Мир, 1991. С. 254.

134. Федорова М.Л., Леденева Т.М. Об исследовании свойства самоподобия трафика мультисервисной сети. *Вестник ВГУ. Серия: Системный анализ и информационные технологии.* 2010. № 1. С. 46-54.

ДОДАТОК А

Документи впровадження основних результатів дисертаційної роботи

Точно засвідчую.
Вчений секретар
спеціалізованої
вченої ради Д64.050.07
Юрій ДОРОФЄЄВ
10.07.2021.

ЗАТВЕРДЖУЮ
Перший проректор
Харківського національного
університету радіоелектроніки
д.т.н., професор _____ І.В. Рубан
«___» _____ 20___ р.

АКТ

про використання в освітньому процесі результатів дисертаційної роботи Булаха Віталія Анатолійовича «Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання», представленої на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології

Комісія у складі:

голови - д.т.н., проф., зав. каф. інфокомунікаційної інженерії ім. В.В. Поповського Лемешка О.В.;

членів - к.т.н., доц., проф. каф. інфокомунікаційної інженерії ім. В.В. Поповського Радівілової Т.А.;

- к.т.н., доц., декана факультету Інфокомунікацій Снігурова А.В.;

розглянула дисертаційну роботу Булаха В.А. та дійшла наступного висновку: матеріали дисертації використовуються в освітньому процесі Харківського національного університету радіоелектроніки, а саме

- методи генерації процесів з заданими фрактальними властивостями;
- використання фрактальних та рекурентних характеристик при класифікації фрактальних процесів методами машинного навчання;
- технологія класифікації упорядкованих масивів даних методами машинного навчання;

що є частиною лекційного курсу та курсу практичних занять з дисципліни «Artificial Intelligence (Штучний інтелект)».

Голова комісії _____ О.В. Лемешко

Члени комісії _____ Т.А. Радівілова

_____ А.В. Снігуров

ЗАТВЕРДЖУЮ
ПрАТ «Фарлеп-Інвест»
_____ А.В. Сіренко
«15» травня 2020 р.

АКТ

про використання результатів дисертаційної роботи Булаха Віталія Анатолійовича за темою «Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання», представлену на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології

Комісія у складі:

голови: Сіренко Андрій Васильович, комерційний керівник;

членів: Мусатов А.П., технічний керівник;

Рощін Е.Д., інженер;

склала даний акт у тому, що результати дисертаційної роботи Булаха В.А., а саме:

- інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями, яка включає збір інформації, навчання методів класифікації, підтримку прийняття рішень для управління вхідними інформаційними потоками з фрактальними властивостями;
- метод виявлення вторгнень в мережі, який базується на використанні бінарної класифікації методами машинного навчання для трафіків даних з фрактальними властивостями;

впроваджено в діяльність підприємства ПрАТ «Фарлеп-Інвест» при розробці практичних рекомендації щодо підвищення рівня безпеки та якості обслуговування в телекомунікаційних мережах в цілому та сприйняття послуг, що надаються кінцевим користувачам.

Голова комісії _____ А.В. Сіренко
(підпис) (ініціали та прізвище)

Член комісії _____ А.П. Мусатов
(підпис) (ініціали та прізвище)

Член комісії _____ Е.Д. Рощін
(підпис) (ініціали та прізвище)

Копію засвідчую.
Вчений секретар
спеціалізованої
вченої ради Д 64.050.07
Юрій ДОРОФЕЄВ
10.02.2021.

Копію засвідчую.
Вчений секретар
спеціалізованої
вченої ради Д64.050.02
Юрій ДОРОФЕЄВ
10.02.2021.

АТ КБ «ПРИВАТБАНК»
Адреса: вул. Грушевського, 1д, м. Київ
01001, Україна
Тел.: +38-056-716-11-31
https://privatbank.ua

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ * НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ "ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ" код 02071180

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційного дослідження

Булаха Віталія Анатолійовича

на тему: «Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання»

Даний акт складено про те, що результати дисертаційної роботи Булаха Віталія Анатолійовича «Інформаційна технологія класифікації впорядкованих масивів даних із фрактальними властивостями методами машинного навчання» впроваджено та використано у діяльності АТ КБ «ПРИВАТБАНК».

Під час підготовки дисертації Булах В.А. розробив інформаційну технологію класифікації впорядкованих масивів даних із фрактальними властивостями. Дана технологія була покладена в основу розробки системи моніторингу комп'ютерної мережі АТ КБ «ПРИВАТБАНК», яка заснована на класифікації вхідних фрактальних інформаційних потоків, що дозволяє підвищити точність та своєчасність виявлення DDoS- атак у мережі.

Результати досліджень, отримані Булахом В.А. під час написання дисертаційної роботи, дозволили зменшити кількість зловмисних утручань в комп'ютерну мережу, витрати часу, трудові та грошові ресурси для моніторингу та обслуговування мережі.

Даний акт не є підставою для проведення взаємних фінансових розрахунків.

Керівник напрямку трудових ресурсів РП І.О. Попов

ДОДАТОК Б

Список публікацій здобувача

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.