

Л.О. Кіріченко, В.А. Булах, М.Ф. Тавалбех, П.П. Зінченко  
**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСИФІКАЦІЇ  
ФРАКТАЛЬНИХ ЧАСОВИХ РЯДІВ**

*Анотація.* В роботі запропоновано інформаційну технологію класифікації часових рядів, що мають фрактальні властивості, на основі методів машинного навчання. Вибір методу класифікації та відповідного набору ознак ґрунтується на мультифрактальних і самоподібних властивостях часових рядів. Як приклад, на основі запропонованої інформаційної технології проведена бінарна класифікація реалізацій нормальних та атакованих трафіків.

*Ключевые слова:* інформаційна технологія, класифікація часових рядів, машинне навчання, фрактальні часові ряди.

**Вступ і мета.** Часові ряди є основною інформацією для розуміння динаміки в складних системах різного типу. Протягом останніх двох десятиліть було запропоновано і розроблено безліч методів інтелектуального аналізу для часових рядів, в тому числі методи машинного навчання [1,2]. Машинне навчання використовуються для різних завдань аналізу часових рядів, зокрема для класифікації.

Багато складних систем мають фрактальну структуру, а їх динаміка представлена часовими рядами, які мають фрактальні (самоподібні) властивості. Аналіз фрактальних властивостей часових рядів широко використовується в різних областях знань. У багатьох випадках виникають проблеми розпізнавання і класифікації фрактальних часових рядів. В останні роки зростає інтерес до методів машинного навчання для аналізу і класифікації фрактальних рядів [3-6].

Однак досі немає єдиного підходу до класифікації часових рядів на основі їх фрактальних властивостей. Мета даної роботи - запропонувати інформаційну технологію класифікації часових рядів, які мають фрактальні властивості, на основі методів машинного навчання.

**Самоподібні та мультифрактальні властивості випадкових процесів.** Самоподібність випадкових процесів полягає в збереженні розподілів ймовірностей при зміні масштабу часу. Стохастичний процес  $X(t)$  є самоподібним з параметром  $H$ , якщо процес  $a^{-H}X(at)$  описується тими ж законами розподілів, що і  $X(t)$ . Параметр  $H$ ,  $0 < H < 1$ , званий показником Херста, являє собою ступінь самоподібності процесу. Поряд з цим показник характеризує міру довгострокової залежності. Якщо  $H > 0.5$ , процес  $X(t)$  є персистентним, тобто трендостійким. У разі  $H < 0.5$  процес має антиперсистентні властивості, а при  $H = 0.5$  значення процесу є незалежними.

Мультифрактальні об'єкти є статистично неоднорідними самоподібними об'єктами та виявляють більш складну скейлінгову поведінку. У в цьому випадку скейлінговою характеристикою є нелінійна функція  $h(q)$  – узагальнений показник Херста. Значення  $h(q)$  при  $q = 2$  збігаються зі значеннями ступеня самоподібності  $H$ . Для монофрактальних процесів узагальнений показник Херста  $h(q) = H$ . Діапазон значень узагальненого показника Херста  $\Delta h(q) = h(q_1) - h(q_2)$  визначає ступінь мультифрактальності: чим більше значення  $\Delta h(q)$ , тим більше виражені мультифрактальні властивості процесу. У разі монофрактальності  $\Delta h(q) = 0$ . Досить зручно для визначення діапазону вибирати позитивні значення параметра  $q_1 = 0.1$  і  $q_2 = 5$ . [7].

**Задача класифікації часових рядів.** Ми можемо сформулювати задачу класифікації часових рядів (ЧР) таким чином. Є множина ЧР, які розділені деяким чином на класи. Визначена скінчена множина ЧР, для яких відомо, до яких класів вони належать. Ця множина є навчальною вибіркою. Класова приналежність інших ЧР не відома. Необхідно побудувати алгоритм, здатний класифікувати довільний ЧР з вихідної множини.

У машинному навчанні існує ряд основних методів класифікації (класифікаторів): дерева прийняття рішень, машини опорних векторів, нейронні мережі та інші. Найчастіше на вхід класифікатора надходить

набір деяких ознак, притаманних даному об'єкту, у нашому випадку - ЧР. На виході ми отримаємо значення класу досліджуваного ЧР.

Одним з найважливіших питань класифікації є вибір ознак, за якими проводиться поділ на класи. Зміна фрактальних властивостей ЧР тягне за собою зміну статистичних та кореляційних властивостей. Тому в якості ознак були обрані фрактальні, статистичні та рекурентні характеристики, розраховані за значеннями ЧР.

Дослідження показали, що статистичними характеристиками, які відображають зміну фрактальних властивостей, є дисперсія, коефіцієнт варіації, медіана, коефіцієнт асиметрії та ін. Як фрактальні ознаки зручно використовувати значення показником Херста та узагальненого показника Херста.

Досить новим підходом до використання ознак ЧР в машинному навчанні є обчислення рекурентних характеристик. Рекурентна діаграма ЧР є масивом точок, де елемент з координатами  $(i, j)$  характеризує близькість точок  $i$  та  $j$  ЧР в фазовому просторі. Чисельний аналіз рекурентних діаграм дозволяє обчислювати кількісні міри складності структур рекурентних діаграм, такі як міра рекурентності, міра детермінізму, міра ентропії і ін. Ці характеристики доцільно застосовувати як ознаки в машинному навчанні.

**Методи класифікації.** Одним з найбільш ефективних методів для вирішення задач класифікації, що виникають в самих різних областях, вважається метод дерев рішень. Він полягає в тому, щоб здійснювати процес розподілу вихідних даних на групи, поки не будуть отримані однорідні їх підмножини. Сукупність правил, які дають таке розбиття, дозволяє потім робити висновок для нових даних. Моделі дерев рішень є нестійкими: навіть невелика зміна в навчальній множині може призвести до істотних змін у структурі дерева. У цьому випадку доцільно використовувати ансамблі моделей.

Бегінг (Bagging) - це технологія класифікації, де все елементарні класифікатори навчаються і працюють незалежно один від одного. Ідея полягає в тому, що класифікатори не виправляють помилки один одного, а компенсують їх при голосуванні. Якщо ансамбль будується на основі

моделей різних типів, то для кожного типу буде свій алгоритм навчання. Ефективність Бегінга досягається завдяки тому, що базові алгоритми, які пройшли навчання за різними підвибірками, виходять досить різними, та їхні помилки взаємно компенсуються при голосуванні. [8].

Випадковий ліс (Random Forest) також є методом Бегінга, але на відміну від його основної версії має кілька особливостей: використовує всередині себе ансамбль тільки регресійних або класифікуючих дерев рішень; в алгоритмі семпліювання крім випадкового вибору навчальних об'єктів, також проводиться випадковий вибір ознак; для кожної підвибірки дерево рішень будується до повного вичерпання навчальних прикладів і не піддається процедурі відсікання гілок [9].

Нейронна мережа. У роботі [10] для проведення класифікації в якості класифікатора використовувався повнозв'язний багат шаровий персептрон. Для запобігання ефекту перенавчання в мережу були включені шари регуляризації, по одному шару після кожного повнозв'язну шару. Як метод регуляризації був використаний метод пакетної нормалізації. Даний метод дозволяє підвищити продуктивність і стабілізувати роботу нейронних мереж. В якості методу навчання був обраний метод стохастичною оптимізації Adam (Adaptive Moment Estimation). Алгоритм оптимізації Adam є розширенням методу стохастичного градієнтного спуску з ітеративним оновленням ваг мережі на основі навчальних даних.

**Інформаційна технологія класифікації фрактальних часових рядів.** Порівняльний аналіз методів класифікації фрактальних часових рядів, представлений в роботах [10-15] дозволяє запропонувати інформаційну технологію класифікації, засновану на методах машинного навчання. Розглянемо її короткий опис, схематично представлений на рис. 1

1. Попередня обробка даних. Під цим пунктом розуміється досить великий спектр алгоритмів, що направлений на виявлення апріорно відомої інформації про природу ЧР, його фрактальних і статистичних властивостей [11].

2. Оцінювання самоподібних і мультифрактальних властивостей ЧР для різних класів. Даний крок є найбільш важливим для вибору методу класифікації. Методи оцінювання  $H$  і  $h(q)$  досить повно описані в [7]. Ключовим моментом є визначення факту, що різні класи ЧР мають досить різні фрактальні властивості. Якщо фрактальні властивості для різних класів практично однакові, то даний підхід не має сенсу використовувати.

Аналіз самоподібних властивостей ЧР 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  $0.45 \leq H \leq 0.55$ . У даному випадку, враховуючи похибки оцінювання показника Херста, ми приймаємо, що діапазон значень  $[0.45, 0.55]$  відповідає дуже малой автокореляційній залежності.

Аналіз мультифрактальних властивостей ЧР, тобто оцінювання узагальненого показника Херста, дозволяє виділити три умовних діапазону  $h(q)$ : слабкі мультифрактальні властивості  $\Delta h(q) < 0.4$ , середні мультифрактальні властивості  $0.4 \leq \Delta h(q) \leq 1$  і сильну мультифрактальність  $\Delta h(q) > 1$ . Дослідження показали, що діапазон мультифрактальних і самоподібних властивостей ЧР відіграє важливе значення для вибору класифікатора і, відповідно, точності класифікації.

3. Дослідження показали, що найбільш добре піддаються класифікації ЧР з сильними мультифрактальними властивостями. В цьому випадку і персистентні і антиперсистентні ряди найбільш точно класифікуються за допомогою випадкового лісу на основі дерев регресії (ВЛ[ДР]) [13,14]. В якості ознак достатньо використовувати статистичні (С) та фрактальні (Ф) характеристики ЧР. Якщо ЧР мають трендостійкість (показник Херста  $H > 0.5$ ), то в якості ознак можна використовувати самі значення ЧР. Це займає набагато більш часу, але не потребує ніякого оцінювання фрактальних характеристик.

4. У разі, коли ЧР явно володіє мультифрактальними властивостями, але діапазон узагальненого показника Херста має деяке невелике значення  $0.4 \leq \Delta h(q) \leq 1$ , в якості ознак також можна застосовувати статистичні та фрактальні характеристики. Причому в персистентному

випадку краще працює метод Бегін з деревами регресії ( Б[ДР] ), а в антиперсистентному - Випадковий Ліс із деревами регресії ( ВЛ[ДР] ).

5. Якщо мультифрактальні властивості ЧР слабо виражені, тобто ЧР можна вважати умовно монофрактальними, то як додатні ознаки треба використовувати рекурентні (Р) характеристики [12]. Дослідження показали, що в разі монофрактальності кращі результати класифікації дає застосування нейронних мереж (НМ) [10].

Найбільш складним випадком класифікації є варіант з класифікацією ЧР, які мають слабку та середню мультифрактальність, та у яких практично відсутня автокореляційна залежність ( $0.45 \leq H \leq 0.55$ ). У цьому випадку пропонується застосувати ансамбль (Бегінг) з декількох операторів, а саме Б{НМ, ВЛ[ДР], Б[ДР]}.

**Приклад застосування інформаційної технології класифікації.** Розглянемо застосування запропонованої технології на прикладі детектування DDoS -атак - хакерських атак на обчислювальну систему з метою створення таких умов, при яких користувачі системи не можуть отримати доступ до системних ресурсів. В даний час DDoS-атаки дозволяють довести до відмови практично будь-яку систему, не залишаючи юридично значимих доказів.

Одним з рішень задачі своєчасного виявлення атаки є розробка класифікатора, який би визначав ймовірність, що приходить трафік є атакою. Останні дослідження показують, що однією з характерних ознак атаки є зміна показника Херста у трафіку, що містить атакуючі файли [16].

Для проведення експериментів були взяті дані трафіків з двох наборів реальних даних [16]. З першого набору були взяті реалізації DDoS-атак, а з другого – дані трафіку реальної мережі Internet Service Provider на каналному рівні. Реалізація трафіку під дією DDoS -атаки є сумою трафіку і реалізації одного з видів атаки. В ході роботи проведені експерименти по обчисленню зміни узагальненого показника Херста під дією DDoS атак.

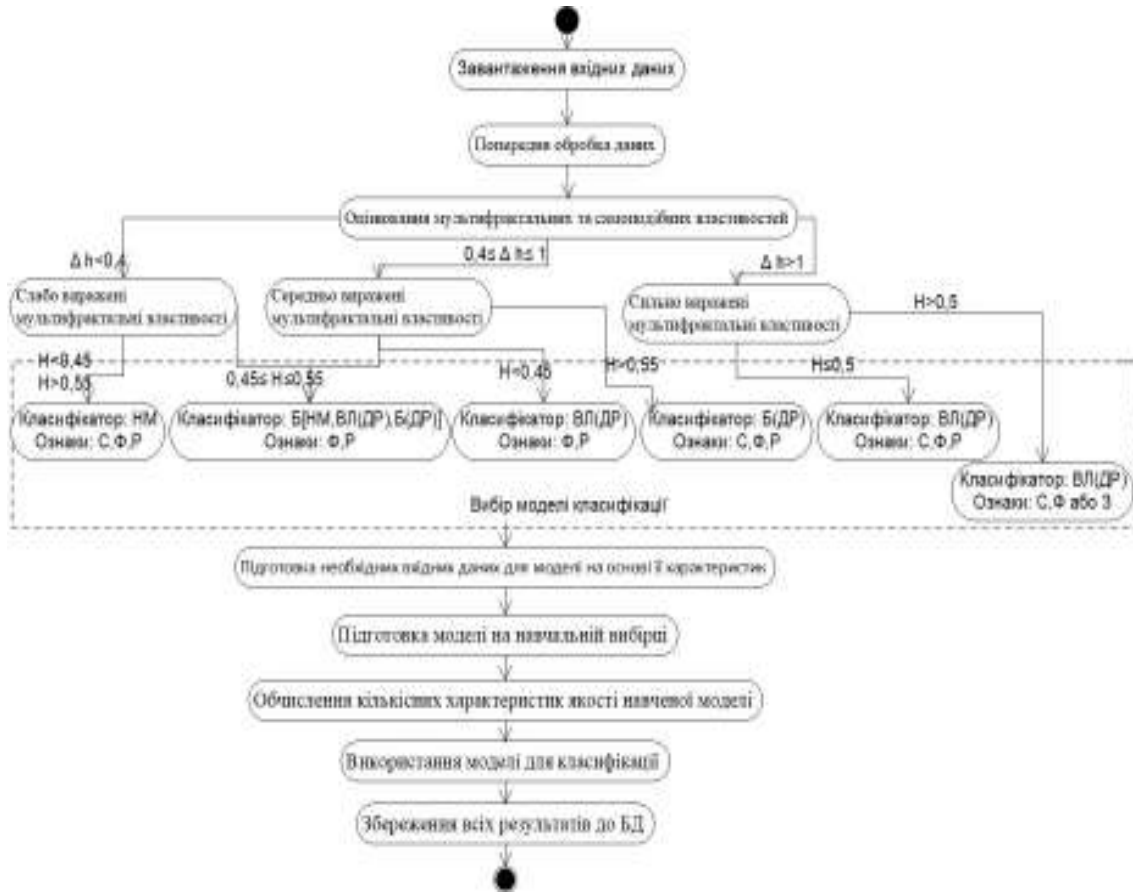


Рисунок 1 - Схема інформаційної технології класифікації часових рядів

У табл.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 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	0.6 1	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 бінарну класифікацію персистентних ЧР з сильно вираженими мультифрактальними властивостями. Методом класифікації був обраний Випадковий ліс на основі регресійних дерев рішень. В якості ознак були використовувані статистичні та фрактальні характеристики реалізацій трафіку. Результатом роботи моделі була ймовірність відповідності реалізації трафіку заданому класу (є атака, чи немає атаки). У табл. 2 представлені результати класифікації для трафіків з різним ступенем самоподібності і атаками, для яких показник Херста був у діапазоні [0.8,0.9].

Таблиця 2

Середні ймовірності визначення атаки

$H$	Ймовірність виявлення атаки	Ймовірність визначення відсутності атаки
[0.5, 0.6]	0.97	0.86
[0.6, 0.7]	0.92	0.84
[0.7, 0.8]	0.88	0.82
[0.8, 0.9]	0.76	0.75

**Висновки.** В роботі розглянуто інформаційну технологію, яка призначена для класифікації часових рядів, що мають фрактальні властивості, на основі методів машинного навчання. В якості методів класифікації запропоновано використовувати такі класифікатори як Бегінг, Випадковий ліс, нейронні мережі. В якості ознак обрані статистичні, фрактальні і рекурентні характеристики часових рядів. Вибір класифікатора і набору ознак ґрунтується на мультифрактальних і самоподібних властивостях досліджуваних часових рядів. Показано, що найбільша точність класифікації досягається для персистентних часових рядів з сильно вираженими мультифрактальними властивостями. Розглянуто приклад бінарної класифікації на основі запропонованої інформаційної технології.



## REFERENCES

1. Esling P., Agon C.: Time series data mining, ACM Computing Surveys, v.46, №.1, (2012).
2. Krisztian Buza: Time Series Classification and its Applications WIMS '18: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics 2018 <https://doi.org/10.1145/3227609.3227690>
3. André L., Coelho V., Clodoaldo A., Lima M.: Assessing fractal dimension methods as feature extractors for EMG signal classification. Engineering Applications of Artificial Intelligence 36 81–98 (2014).
4. Symeonidis S.: Sentiment analysis via fractal dimension. Proceedings of the 6th Symposium on Future Directions in Information Access, 48-50 (2015).
5. S. P. Arjunan, D. K. Kumar, G. R. Naik 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, 4821-4 (2010).
6. Ledesma-Orozco, S.E. Ruiz, 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, pp. 480-485 (2011). doi: 10.1109/ICNSC.2011.5874873
7. Kirichenko L., Radivilova T. Estimation of multifractal characteristics of stochastic processes. International Journal Information content and processing. 2018. Vol. 5 (2). P. 106–141.
8. Breiman L. Bagging predictors. Machine Learning. (1996), 24 (2), P.123–140.
9. Breiman L. Random Forests. Machine Learning. (2001), 45 (1), P.5–32.
10. Kirichenko L., Radivilova T., Bulakh V. Binary Classification of Fractal Time Series by Machine Learning Methods. Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing, vol 1020. Springer, Cham, 2018. P. 701–711. doi: [https://doi.org/10.1007/978-3-030-26474-1\\_49](https://doi.org/10.1007/978-3-030-26474-1_49).
11. Kirichenko L., Radivilova T., Bulakh V. Generalized approach to Hurst exponent estimating by time series. Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska. 2018. Volume 8, No. 1. P. 28–31. doi: 10.5604/01.3001.0010.8639.

12. Kirichenko L., Radivilova T., Bulakh V. Classification of Fractal Time Series Using Recurrence Plots. Problems of Infocommunications. Science and Technology (PIC S&T): Proceedings of International Science-Practical Conference, 2018, Kharkiv, Ukraine. IEEE, 2018. P. 719–724. doi: 10.1109/INFOCOMMST.2018.8632010.
13. 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. 2105. P. 457–460.
14. Bulakh V., Kirichenko L., Radivilova T. Time Series Classification Based on Fractal Properties. Data Stream Mining & Processing (DSMP): 2018 IEEE Second International Conference, Lviv, Ukraine. IEEE, 2018. P. 198–201. doi: 10.1109/DSMP.2018.8478532.
15. Kirichenko L., Radivilova T., Bulakh V. Machine Learning in Classification Time Series with Fractal Properties. Data. 2019. Vol. 4, issue 1, 5. P. 1–13. doi:10.3390/data4010005 8632010.

Received 03.02.2020.

Accepted 10.02.2020.

### **Информационная технология классификации фрактальных временных рядов**

*Предложена информационная технология классификации фрактальных временных рядов на основе методов машинного обучения. Выбор метода классификации и соответствующего набора признаков основывается на мультифрактальных и самоподобных свойствах временных рядов. Используя предложенную технологию, рассмотрен пример бинарной классификации временных рядов.*

### **Information technology of classification of fractal time series**

*We can formulate the task of time series classification in this way. There are many time series that are divided into classes in some way. A finite set of time series is defined for which it is known to which classes they belong. This set is a training set. Belonging to the class of other time series is not known. It is necessary to get an algorithm of classifying an arbitrary time series from the initial set.*

*In machine learning, there are a number of basic classification methods: decision trees, support vector machines, neural networks, and others. Most often, a set of some features of given object, in our case, time series, comes to the input of the classifier. At the output, we get the class value of the investigated time series.*

*One of the most important classification issues is the selection of features by which the division into classes is carried out. A change in the fractal properties of time series entails a change in statistical and correlation properties. Therefore, fractal, statistical and recurrent characteristics calculated from time series were chosen as features.*

*Studies have shown that the statistical characteristics that reflect changes in fractal properties are variance, coefficient of variation, median, asymmetry coefficient, etc. As fractal*

features, it is convenient to use the value of the Hurst exponent and the generalized Hurst exponent.

A fairly new approach to the use of time series features in machine learning is the calculation of recurrence characteristics. A recurrence plot is an array of points, where an element with coordinates  $(i, j)$  characterizes the proximity of points  $i$  and  $j$  of the time series in the phase space. The numerical analysis of recurrence plots allows one to calculate the quantitative degrees of complexity of the structures of recurrence plots, such as a measure of recurrence, a measure of determinism, a measure of entropy, etc. These characteristics are useful as features in machine learning

A comparative analysis of the classification methods for fractal time series allows us to offer an information classification technology based on machine learning methods. Let's consider its short description.

1. Pre-processing of data.

2. Evaluation of self-similar and multifractal properties of time series for different classes. The key point is the determination of the fact that different classes of time series have quite different fractal properties. If the fractal properties for different classes are almost the same, then this approach does not make sense to use.

3. Studies have shown that the classification of time series with strong multifractal properties is the best. In this case, the series are most accurately classified using a random forest based on regression trees. It is sufficient to use the statistical and fractal characteristics as features.

4. In the case when the time series has average multifractal properties, statistical and fractal characteristics can be used as features. In the persistent case, the bagging method with regression trees is used, and in the antipersistent case the Random Forest with regression trees is used.

5. If the multifractal properties are weakly expressed, that is, the time series can be considered as conditionally monofractal, then recurrent characteristics should be used as additional features. Studies have shown that in the case of monofractality, the best classification results are obtained by the use of neural networks.

**Кириченко Людмила Олеговна** - д.т.н., професор кафедри прикладної математики Харківського національного університету радіоелектроніки.

**Булах Віталій Анатольевич** - асистент кафедри системотехніки Харківського національного університету радіоелектроніки.

**Зинченко Петр Петрович** - аспірант кафедри прикладної математики Харківського національного університету радіоелектроніки.

**Тавалбех Максим Хаджем** - аспірант кафедри прикладної математики Харківського національного університету радіоелектроніки

**Кіріченко Людмила Олегівна** – д.т.н., професор кафедри прикладної математики Харківського національного університету радіоелектроніки.

**Булах Віталій Анатолійович** – асистент кафедри системотехніки Харківського національного університету радіоелектроніки.

**Зінченко Петро Петрович** – аспірант кафедри прикладної математики Харківського національного університету радіоелектроніки.

**Тавалбех Максим Хаджем** – аспірант кафедри прикладної математики Харківського національного університету радіоелектроніки

**Kirichenko Lyudmyla** - Doctor of Technical Sciences, Professor, Department of Applied Mathematics, Kharkov National University of Radio Electronics.

**Bulakh Vitalii** - Assistant of the Department of Systems Engineering, Kharkov National University of Radio Electronics.

**Zinchenko Petro** - graduate student of the Department of Applied Mathematics, dozens of Kharkov National University of Radio Electronics.

**Tawalbeh Maxim** - graduate student of the Department of Applied Mathematics, Kharkov National University of Radio Electronics.