

ТЕХНІЧНІ НАУКИ

DOI: <https://doi.org/10.33216/1998-7927-2022-271-1-5-10>

УДК: 004.852

ДОСЛІДЖЕННЯ МЕТОДІВ АНАЛІЗУ НЕЛІНІЙНОЇ
ДИНАМІКИ ЧАСОВИХ РЯДІВ

Білобородова Т.О.

TIME SERIES ANALYSIS USING NONLINEAR DYNAMICS METHODS

Biloborodova T.O.

Важливою складовою аналізу даних є вилучення значущих ознак про систему, використовуючи дані часових рядів. Більшість методів традиційного аналізу, зосереджена на даних часових рядів або частотної області. Аналіз нелінійної динаміки ґрунтується на представленні системи у просторі станів, відображаючи поведінку в n -мірному просторі, осі якого є змінними станами. Традиційні методи дослідження діагностичних ознак, в умовах нестационарності даних, не завжди забезпечують необхідну достовірність результатів діагностики фізіологічних процесів. Це зумовлює використання додаткових методів, заснованих на використанні порядкової мережі, та направлених на поділ часових рядів на короткі проміжки часу, протягом яких фізіологічна система знаходиться в стаціонарних умовах. Ентропії є мірами оцінки складності порядкової мережі та означають міру складності, хаотичності чи невизначеності системи.

Проведено дослідження з застосування порядкової мережі та мір оцінки її складності для визначення хаотичних процесів в часових рядах даних фізіологічних процесів, що можуть вказувати на патологічні відхилення досліджуваних процесів. Формалізовано процес перетворення часових рядів у порядкову мережу. Виконано реалізацію перетворення часових рядів QRS-комплексів ЕКГ у порядкову мережу та розрахунок мір складності: умовної ентропії перестановки та ентропії глобального вузла. Проведений розвідувальний аналіз даних отриманих результатів умовної ентропії перестановки та ентропії глобального вузла на основі візуалізації кореляцій в даних дозволив якісно оцінити відмінність умовної ентропії перестановки та ентропії глобального вузла для двох типів зразків QRS-комплексів ЕКГ. Статистична значущість, визначена з використанням t -критерій Стьюдента та значення r -критерію менше 0,05 свідчить, що різниця дисперсії мір для двох типів зразків QRS-комплексів ЕКГ є статистично значущою.

За результатами проведеного дослідження зроблено висновки про доцільність застосування порядкової мережі

та мір оцінки її складності для визначення хаотичних процесів в клінічних даних для диференційного аналізу патологічних та нормальних фізіологічних процесів.

Ключові слова: аналіз часових рядів, нелінійна динаміка, порядкова мережа, умовна ентропія перестановки, ентропія глобального вузла

Вступ. Інтелектуальний аналіз даних часто включає завдання вилучення достовірної інформації про складну систему на основі даних дані часових рядів. Більшість клінічних даних може бути представлена часовими рядами, що характеризують фізіологічні процеси систем організму людини. Клінічні дані можуть бути представлені лінійними або нелінійними фізіологічними процесами. Лінійна (стаціонарна), інваріантна у часі система може бути досить добре описана традиційними моделями інтелектуального аналізу даних. У лінійних системах часто можна з упевненістю припустити і легко визначити патологічні процеси, що, в даному випадку, виходять за межі встановлених значень, що відповідають нормальному фізіологічному стану. Нестационарні або нелінійні процеси представляють додаткову інформацію про стан організму, проте вимагають нестандартних підходів для отримання корисної інформації.

Більшість методів традиційного аналізу, зосереджена на даних часових рядів або частотної області. Аналіз нелінійної динаміки ґрунтується на представленні процесу у просторі станів, відображаючи поведінку в n -мірному просторі, осі якого є змінними станами. Електроенцефалограма, електрокардіограма, електроміограма тощо являють собою часовий ряд, що описує зміни в часі сумарного електричного потенціалу досліджуваної системи організму. Ці дані мають нелінійні властивості та складну по-

ведінку [1,2,3]. Оцінка їх рівня регулярності може бути корисною в багатьох додатках [4,5]. Складність подібних часових рядів зазвичай оцінюється виходячи з припущення, що дані створені стаціонарною динамічною системою, але можуть розвивати нерегулярну поведінку через свою нелінійність. Однак клінічні дані фізіологічних процесів систем організму людини часто є нестаціонарними і традиційні діагностичні ознаки не завжди забезпечують необхідну достовірність результатів діагностики фізіологічних процесів [6]. Це зумовлює використання додаткових методів, направлених на поділ часових рядів на короткі проміжки часу, протягом яких фізіологічна система знаходиться в стаціонарних умовах. Аналіз нелінійної динаміки часових рядів клінічних даних заснований на переході від їхнього скалярного представлення, що реєструється в часовій області, до векторного представлення у фазовому просторі. Використання порядкової мережі [7] для аналізу нелінійної динаміки дає змогу отримати корисну інформацію про зміни фізіологічних процесів систем організму. Методи на основі порядкової мережі та застосування інтервалу затримки не тільки дають корисну інформацію про поведінку незмінних змінних, але й дають деяке уявлення про кількість незалежних змінних станів, що фактично міститься у часових рядах даних фізіологічних систем. Для оцінки порядкової мережі використовуються постійні заходи, які залишаються постійними незалежно від методу спостереження системи.

В якості незмінних критеріїв для аналізу клінічних даних фізіологічних процесів у дослідженнях використовують фазову ентропію перестановок Ренні [8], складність Лемпеля-Зіва та показник Ляпунова [9] та ентропію Колмогорова-Сіная [10].

Ентропії є мірами складності часових рядів і означають міру складності, хаотичності чи невизначеності системи: що менше елементи системи підпорядковані якомусь порядку, то вище ентропія. Окрім ентропії Колмогорова-Сіная, в дослідженнях використовуються ентропія перестановки (англ. permutation entropy (PE)), умовна ентропія перестановки (англ. conditional permutation entropy (CPE)), ентропія глобального вузла (англ. global node entropy (GNE)) [7]. Дослідження [11] показало зміну ступеня мір складності фізіологічних процесів даних у разі виникнення патології. Таким чином, ці міри при аналізі клінічних даних характеризують хаотичність механізмів фізіологічних процесів і можуть використовуватись для підвищення достовірності оцінки характеру діяльності систем організму.

Метою роботи є дослідження застосування порядкової мережі та мір оцінки її складності для визначення хаотичних процесів в клінічних даних фізіологічних процесів, що можуть вказувати на патологічні відхилення досліджуваних процесів.

Для вирішення поставленої мети завдання дослідження визначені наступним чином: (1) формалізація процесу перетворення часових рядів у порядкову мережу, (2) реалізація процесу перетворення

часових рядів у порядкову мережу та розрахунок мір її складності, (3) розвідувальний аналіз даних отриманих результатів мір складності, (4) визначення статистичної значущості отриманих результатів диференціації патологічних та нормальних фізіологічних процесів.

Перетворення часових рядів на порядкову мережу. Відображення часового ряду в порядковій мережі базується на його порядковому розбитті з урахуванням послідовних можливих шаблонів даних [7]. Порядкове розбиття параметризується вкладеною розмірністю вибірки m та інтервалом затримки τ . Визначимо часовий ряд клінічних даних фізіологічного процесу системи організму людини як серію рівновіддалених спостережень $\mathbf{x} = \{x_1, \dots, x_m, \dots, x_M\}$, де m ціле число та представляє розмірність вкладеної вибірки часового ряду, M – загальна кількість часових кроків в часовому ряді, що розбивається. Отримання вкладених векторів $\mathbf{z} = \{z_1, z_2, \dots, z_{M-(m-1)\tau}\}$ проводиться на основі зв'язків розмірності m та затримки τ .

Кожен вектор $z_i = \{x_1, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}\}$ відображається в послідовність символів відповідно до значення амплітуди кожного елемента в ньому. Якщо два елементи z_i мають однакове значення ($x_i = x_j$), символ присвоюється у порядку появи. Унікальний набір всіх послідовностей символів $\mathbf{s}$ представлений часовим рядом $\mathbf{s} = \{s_1, s_2, \dots, s_{M-(m-1)\tau}\}$, де кожен $s_i \in \mathbf{s}$ представляє собою вузол для побудови порядкової мережі.

Міри визначення складності для оцінки хаотичності часових рядів. Наукова гіпотеза базується на припущенні, що використання мір складності для оцінки хаотичності порядкової мережі дозволяє визначити достовірно значуще розрізнення відхилень у динаміці часових рядів. Міри складності розраховуються на основі РЕ часового ряду, яка відповідає ентропії Шеннона відповідного набору порядкових символів $\mathbf{s}$.

$$PE = - \sum_i p_i \log p_i, \quad (1)$$

де p_i представляє собою функцію маси ймовірності $P(S=s_i)$ для $s_i \in \mathbf{s}$ та оцінюється шляхом підрахунку відносної появи кожного символу у символній динаміці S . РЕ детально розглянута в [12].

CPE запропонована в дослідженні [13] та розраховується наступним чином (2).

$$CPE = \sum_i (-p_i \sum_j p_{i,j} \log p_{i,j}) \quad (2)$$

де p_i використовується з рівняння (1), а $p_{i,j}$ – елемент стохастичної матриці P і є ймовірністю переходу від s_i до s_j , що оцінюється на основі символної динаміки S .

GNE розраховується на основі елементів $p_{i,j}^T$ модифікованої стохастичної матриці P^T порядкової мережі шляхом усереднення мережі на основі стаціонарного розподілу (3).

$$GNE = \sum_i p_i (-\sum_j p_{ij}^T \log p_{ij}^T), \quad (3)$$

Перетворення часових рядів у порядкову мережу та розрахунок мір її складності на даних ЕКГ. Реалізація перетворення часових рядів ЕКГ та аналізу отриманої порядкової мережі з використанням мір складності CPE, GNE для диференціації нормальних QRS комплексів від патологічних проведена з використанням набору даних ЕКГ [14]. Досліджуваний набір даних містить часові ряди зразків QRS комплексів, виділених з ЕКГ. Кожен зразок QRS комплексу містить 140 часових кроків. Кожен зразок відповідає одному серцебиттю, тобто QRS-комплексу від пацієнта із застійною серцевою недостатністю. В цьому дослідженні використано 2 типи QRS-комплексів: тип 1 - нормальні зразки та тип 2 - зразки з наявністю ознак передчасного скорочення шлуночків. Сформований набір даних включає по 500 зразків кожного з цих двох станів. Для ЕКГ більш довгі вкладені вектори можуть визначати QRS-комплекси ЕКГ, а короткі – сегменти QRS-комплексу, пов'язані з функціональними компонентами: хвилями або сегментами комплексу. Для оцінки окремих компонентів QRS-комплексу з порядкового розбиття CPE, GNE кожного часового ряду обчислюється для короткого інтервалу затримки. Значення розмірності вкладеного вектору встановлено на $m = 5$, значення інтервалу затримки τ встановлено на рівні $\tau = 8$. Для кожного часового ряду ЕКГ набору даних розраховані міри складності CPE, GNE.

Розвідувальний аналіз даних отриманих результатів мір складності. Розвідувальний аналіз даних направлений на визначення наявності або відсутності кореляції в даних отриманих мір складності між двома типами описаних вище зразків.

На першому етапі розвідувального аналізу даних отриманих мір складності побудовано діаграму розсіювання кореляційного зв'язку між отриманими оцінками мір складності CPE, GNE і типами зразків представлена на рис. 1.

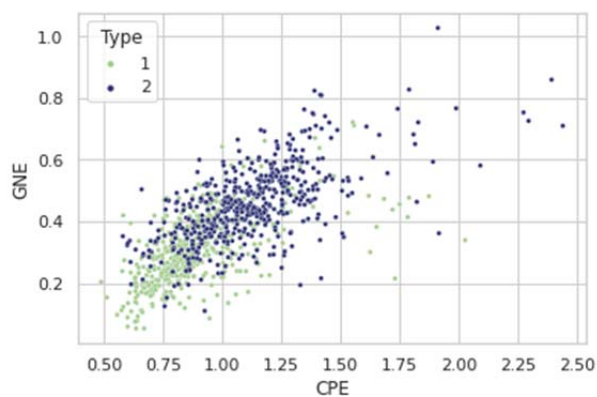


Рис. 1. Діаграма розсіювання кореляційного зв'язку між CPE, GNE та досліджуваними типами зразків

Дослідження розподілу значень мір складності CPE, GNE для двох типів зразків даних проведено з застосуванням гістограм, що представлені на рис. 2, 3.

Як можна побачити з гістограм розподілу даних зразків типу 1, більша частина значень отриманої міри складності CPE лежить у межах від 0.75 до 1, а для зразків типу 2 у межах від 1 до 1.25. Більша частина значень отриманої міри складності GNE лежить у межах від 0.2 до 0.4, а для зразків типу 2 у межах від 0.3 до 0.5. Отримані гістограми дозволяють оцінити відмінність типів зразків за результатами кількісних показників мір складності CPE, GNE.

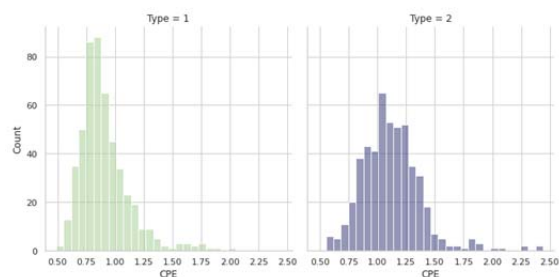


Рис. 2. Гістограми розподілу значень CPE для двох типів зразків

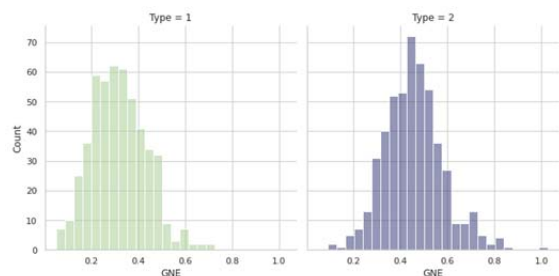


Рис. 3. Гістограми розподілу значень GNE для двох типів зразків

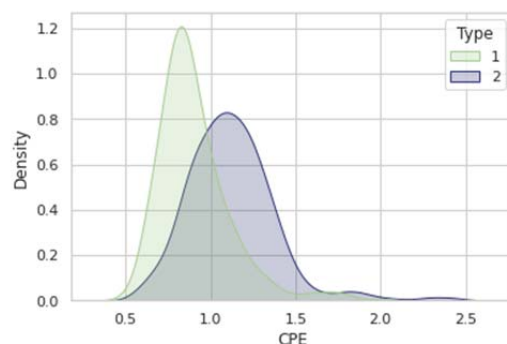


Рис. 4. Графік оцінки щільності даних CPE у відповідності до типу зразків

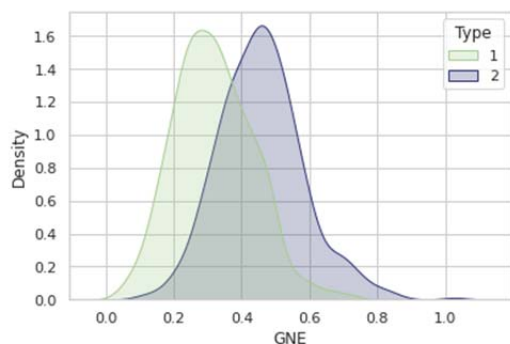


Рис. 5. Графік оцінки щільності даних GNE у відповідності до типу зразків

Графіки оцінки щільності даних, представлені на рис. 4, 5 візуалізують розподіл отриманих CPE, GNE відповідно.

Графіки оцінки щільності даних представляють дані, використовуючи безперервну криву щільності ймовірностей за якими можна зробити висновок про розподіл значень оцінок між типами зразків QRS-комплексів. Області перекриття значень оцінок представлені третім кольором. Таким чином, можна судити про наявність великої кількості значень CPE та GNE, які притаманні певному типу зразків і, таким чином, дають певний рівень достовірності при їх диференціюванні.

Визначення статистичної значущості. Для визначення статистичної значущості відмінності двох типів досліджуваних QRS-комплексів на основі мір складності CPE та GNE використано t-критерій Стюдента. Для перевірки нульової гіпотези про рівність дисперсій отриманих оцінок двох типів зразків QRS-комплексів значення p -критерію нижче 0,05 вважалося статистично значущим, з довірчим інтервалом 95%. Для отриманих значень CPE, GNE отримано значення p -критерію менше 0,05. Таким чином, наявна різниця дисперсій вважається статистично значущою, нульова гіпотеза про рівність дисперсій відкидається.

Таким чином, підтверджено статистичну значущість відмінності зразків типу 1 з нормальним QRS-комплексом від зразків типу 2 – з наявністю ознак передчасного скорочення шлуночків на підставі мір складності отриманих в результаті оцінки складності порядкових мереж досліджуваних зразків.

Висновки. Аналіз динаміки клінічних даних фізіологічних процесів людини дозволяє виділити додаткові знання з досліджуваних даних. Нестационарність деяких фізіологічних процесів вимагає використання додаткових методів, таких як порядкова мережа, для проведення аналізу таких даних. Для оцінки порядкової мережі на основі даних фізіологічних процесів використовуються міри складності, такі як умовна ентропія перестановки та ентропія глобального вузла. Міри складності, у випадку ана-

лізу клінічних даних функціональних систем організму, вказують на хаотичність цих даних та можуть свідчити про патологічні зміни досліджуваних систем організму.

В результаті проведеного дослідження з застосування порядкової мережі та мір оцінки її складності для визначення хаотичних процесів в клінічних даних фізіологічних процесів формалізовано процес перетворення часових рядів у порядкову мережу. Виконано реалізацію перетворення часових рядів QRS-комплексів ЕКГ у порядкову мережу та розрахунок мір складності: умовної ентропії перестановки та ентропії глобального вузла. Проведений розвідувальний аналіз даних отриманих результатів умовної ентропії перестановки та ентропії глобального вузла на основі візуалізації кореляцій в даних дозволив якісно оцінити відмінність умовної ентропії перестановки та ентропії глобального вузла для двох типів зразків QRS-комплексів ЕКГ. Статистична значущість, визначена з використанням t -критерію Стюдента та значення p -критерію менше 0,05 свідчить, що різниця дисперсій мір для двох типів зразків QRS-комплексів ЕКГ є статистично значущою.

Таким чином, зроблено висновок про доцільність застосування порядкової мережі та мір оцінки її складності для визначення хаотичних процесів в клінічних даних для диференційного аналізу патологічних та нормальних фізіологічних процесів.

Література

1. Liu, T., Yao, W., Wu, M., Shi, Z., Wang, J., Ning, X. Multiscale permutation entropy analysis of electrocardiogram. *Physica A: Statistical Mechanics and its Applications*. – 2017. – Т. 471. – С. 492-498.
2. Yu, H., Liu, J., Cai, L., Wang, J., Cao, Y., Hao, C. Functional brain networks in healthy subjects under acupuncture stimulation: an EEG study based on nonlinear synchronization likelihood analysis. *Physica A: Statistical Mechanics and its Applications*. – 2017. – Т. 468. – С. 566-577.
3. Mesin, L., Dardanello, D., Rainoldi, A., Boccia, G. Motor unit firing rates and synchronisation affect the fractal dimension of simulated surface electromyogram during isometric/isotonic contraction of vastus lateralis muscle. *Medical engineering & physics*. – 2016. – Т. 38. – №. 12. – С. 1530-1533.
4. Seely A. J. E., Christou N. V. Multiple organ dysfunction syndrome: exploring the paradigm of complex nonlinear systems. *Critical care medicine*. – 2000. – Т. 28. – №. 7. – С. 2193-2200.
5. Gao, Z. K., Fang, P. C., Ding, M. S., Jin, N. D. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Experimental Thermal and Fluid Science*. – 2015. – Т. 60. – С. 157-164.
6. Файнзильберг Л.С. Компьютерная диагностика по фазовому портрету электрокардиограммы / Л.С. Файнзильберг. – К. : Освіта України, 2013. – 191 с.

7. McCullough, M., Small, M., Iu, H. H. C., Stemler, T. Multiscale ordinal network analysis of human cardiac dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* – 2017. – T. 375. – №. 2096. – C. 20160292.
8. Kang H., Zhang X. The influence of parameter selection for Rényi phase permutation entropy on abnormal change detection. 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP). – IEEE, 2019. – C. 1-6.
9. Sun, Q., Wang, Q., Ji, B., Wu, W., Huang, W., Wang, C. The Cardiodynamicsgram Based Early Detection of Myocardial Ischemia Using the Lempel-Ziv Complexity. *IEEE Access.* – 2020. – T. 8. – C. 207894-207904.
10. Scarciglia, A., Catrambone, V., Bonanno, C., Valenza, G. Quantifying partition-based Kolmogorov-Sinai Entropy on Heart Rate Variability: a young vs. elderly study. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). – IEEE, 2021. – C. 5469-5472.
11. Майоров О. Ю., Фенченко В. Н. Повышение надежности исследований детерминированного хаоса в биоэлектрической активности (ЭЭГ, ЭКГ и вариабельности сердечного ритма) методами нелинейного анализа. *Клиническая информатика и телемедицина.* – 2009. – №. 5, Вып. 6. – С. 10-17.
12. Bandt C., Pompe B. Permutation entropy: a natural complexity measure for time series. *Physical review letters.* – 2002. – T. 88. – №. 17. – C. 174102.
13. Unakafov A. M., Keller K. Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena.* – 2014. – T. 269. – C. 94-102.
14. Chen Y., Hao, Y., Rakthanmanon T., Zakaria J., Hu B., Keogh E. A general framework for never-ending learning from time series streams. *Data mining and knowledge discovery.* – 2015. – T. 29. – №. 6. – C. 1622-1664.
6. Fainzilberg L.S. *Electrocardiogram phase portraits in computer diagnostics.* Kiev: Osvita Ukrainy, 2013. 191 p.
7. McCullough, M., Small, M., Iu, H. H. C., Stemler, T. Multiscale ordinal network analysis of human cardiac dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* – 2017. – T. 375. – №. 2096. – P. 20160292.
8. Kang H., Zhang X. The influence of parameter selection for Rényi phase permutation entropy on abnormal change detection. 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP). – IEEE, 2019. – P. 1-6.
9. Sun, Q., Wang, Q., Ji, B., Wu, W., Huang, W., Wang, C. The Cardiodynamicsgram Based Early Detection of Myocardial Ischemia Using the Lempel-Ziv Complexity. *IEEE Access.* – 2020. – T. 8. – P. 207894-207904.
10. Scarciglia, A., Catrambone, V., Bonanno, C., Valenza, G. Quantifying partition-based Kolmogorov-Sinai Entropy on Heart Rate Variability: a young vs. elderly study. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). – IEEE, 2021. – P. 5469-5472.
11. Mayorov, O., Fenchenko V. Increase reliability of bioelectric activity (EEG, ECG and HRV) deterministic chaos researches by the nonlinear analysis methods. *J. Klin.inform.telemed.* – 2009. – №. 5, Vol. 6. – P. 10-17.
12. Bandt C., Pompe B. Permutation entropy: a natural complexity measure for time series. *Physical review letters.* – 2002. – T. 88. – №. 17. – P. 174102.
13. Unakafov A. M., Keller K. Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena.* – 2014. – T. 269. – P. 94-102.
14. Chen Y., Hao, Y., Rakthanmanon T., Zakaria J., Hu B., Keogh E. A general framework for never-ending learning from time series streams. *Data mining and knowledge discovery.* – 2015. – T. 29. – №. 6. – P. 1622-1664.

References

1. Liu, T., Yao, W., Wu, M., Shi, Z., Wang, J., Ning, X. Multiscale permutation entropy analysis of electrocardiogram. *Physica A: Statistical Mechanics and its Applications.* – 2017. – T. 471. – P. 492-498.
2. Yu, H., Liu, J., Cai, L., Wang, J., Cao, Y., Hao, C. Functional brain networks in healthy subjects under acupuncture stimulation: an EEG study based on nonlinear synchronization likelihood analysis. *Physica A: Statistical Mechanics and its Applications.* – 2017. – T. 468. – P. 566-577.
3. Mesin, L., Dardanella, D., Rainoldi, A., Boccia, G. Motor unit firing rates and synchronisation affect the fractal dimension of simulated surface electromyogram during isometric/isotonic contraction of vastus lateralis muscle. *Medical engineering & physics.* – 2016. – T. 38. – №. 12. – P. 1530-1533.
4. Seely A. J. E., Christou N. V. Multiple organ dysfunction syndrome: exploring the paradigm of complex nonlinear systems. *Critical care medicine.* – 2000. – T. 28. – №. 7. – P. 2193-2200.
5. Gao, Z. K., Fang, P. C., Ding, M. S., Jin, N. D. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Experimental Thermal and Fluid Science.* – 2015. – T. 60. – P. 157-164.

Biloborodova T.O. Time series analysis using nonlinear dynamics methods

An important component of data mining is the extraction of significant system features using time series data. Most traditional analysis methods focus on time series or frequency domain data. The analysis of nonlinear dynamics is based on the representation of the system in the state space, reflecting the behavior in the n-dimensional space, the axes of which are variable states. Traditional methods of studying diagnostic signs in non-stationary data do not always provide the necessary confidence of diagnosing physiological processes. This leads to the use of additional methods based on the use of an ordinal network and aimed at partition time series into short time intervals during which the physiological system is in stationary conditions. Entropies are complexity measures of ordinal network and indicate the complexity, randomness or uncertainty of the system.

The study was carried out on the implementation of the ordinal network and complexity measures to determine chaotic processes in the time series of these physiological processes, which may indicate pathology processes. The process of transforming time series into an ordinal network is formalized. The implementation of the transformation of the time series of ECG QRS complexes into an ordinal network and the calculation of complexity measures: the conditional permutation entropy and the global node entropy have been

performed. The exploratory data analysis of the obtained results of the conditional permutation entropy and the global node entropy based on the visualization of data correlations made it possible to qualitatively evaluate the difference in the conditional permutation entropy and the global node entropy for two types of ECG QRS samples. Statistical significance, determined using the Student's *t*-test and a *p*-test value of less than 0.05, indicates that the difference in measure variance for the two types of ECG QRS samples is statistically significant.

Based on the results of the study, a conclusion was made about the feasibility of using an ordinal network and measures for assessing its complexity to determine chaotic processes in clinical data for differential analysis of pathological and normal physiological processes.

Keywords: time series analysis, nonlinear dynamics, ordinal network, conditional permutation entropy, global node entropy

Білобородова Тетяна Олександрівна – к.т.н., доцент, докторант Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова, e-mail: beloborodova.t@gmail.com

Стаття подана 15.01.2022.