

ГОЛОБОРОДЬКО Владислав

Східноукраїнський національний університет імені Володимира Даля

<https://orcid.org/0009-0007-2385-9730>

e-mail: asp-126-24-3@snu.edu.ua

ЛИФАР Володимир

Східноукраїнський національний університет імені Володимира Даля

<https://orcid.org/0000-0002-7860-9663>

e-mail: lifar@snu.edu.ua

МЕТОД КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ІДЕНТИФІКАЦІЇ ОСОБИСТОСТІ НА ОСНОВІ ДАНИХ ЕЕГ

У статті розглянуто метод ідентифікації особистості на основі аналізу електроенцефалограм (ЕЕГ). Запропоновано метод кластерного аналізу багатовимірних даних електроенцефалограм для визначення індивідуальних характеристик мозкової активності.

Проведено експериментальні дослідження, що підтвердили можливість класифікації енцефалограм та визначення їхньої приналежності до конкретного суб'єкта.

Результатом класифікації є віднесення контрольної електроенцефалограми до відповідного класу, що відповідає конкретній особі. Запропонований метод продемонстрував високу достовірність.

Ключові слова: електроенцефалограма, кластерний аналіз, біометрична ідентифікація, статистичний аналіз, сигнатура барицентра, емоційний стан, стан спокою

HOLOBORODKO Vladyslav, LYFAR Volodymyr

East Ukrainian National University

CLUSTER ANALYSIS METHOD FOR PERSONALITY IDENTIFICATION BASED ON EEG DATA

In this work, we propose a method for personality identification based on the analysis of electroencephalographic (EEG) data using cluster analysis techniques. The method is grounded on the hypothesis that EEG signals contain biometric information specific to each individual. Our approach involves capturing EEG data under controlled conditions of rest and minimal external interference, allowing us to extract unique patterns from each subject's brain activity.

The experimental setup utilizes a 22-channel EEG recording device, with repeated signal acquisitions taken from two test subjects under identical conditions. Each subject undergoes multiple testing sessions with repeated electrode placements to ensure signal variability and consistency. From the EEG recordings, we extract amplitude tables, power spectrum data, and frequency distributions, which serve as the foundation for statistical and cluster analysis.

The clustering algorithm centers around the calculation of barycenters for each EEG channel. These barycenters represent average amplitude values and serve as a statistical signature of a subject's EEG profile. We use the Euclidean distance to compare new (control) EEG signals to these established clusters. A smaller distance indicates a higher likelihood of match, confirming the EEG's identity correspondence.

Data processing and analysis are carried out using tools such as MathCad and Excel, allowing for efficient manipulation of EEG matrices and extraction of meaningful metrics. Clusters of amplitude, power, and frequency are formed for each test subject. The verification process involves comparing control EEG signals to these clusters, evaluating how closely the new data aligns with the established barycenters.

Results demonstrate that the control EEG of subject "Andre" deviates 49% from their barycenter, while it deviates 54% from "Vlad's" barycenter, confirming Andre's identity. Similarly, Vlad's EEG is closer to his own barycenter than to Andre's. The same identification accuracy is observed across power and frequency clusters.

This cluster-based approach proves to be a reliable method for biometric identification using EEG. The presented method does not require complex preprocessing or machine learning models, relying instead on statistical consistency and barycenter proximity. Our findings support the viability of EEG data as a biometric identifier and highlight the potential for real-world application in identity verification systems.

Keywords: electroencephalogram, cluster analysis, biometric identification, statistical analysis, barycenter signature, emotional state, resting state

Стаття надійшла до редакції / Received 16.10.2025

Прийнята до друку / Accepted 29.11.2025

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Дослідження електричної активності мозку є одним із ключових напрямів у нейронауці[1]. Електроенцефалограма людини містить глибоко зашифровану інформацію про діяльність мозку та відображає електромагнітну активність, що містить внутрішній відгук на поточний вплив на людський організм ззовні[2].

Дістати цю інформацію в деякому структурованому вигляді практично неможливо. Це пов'язано з тим, що фізичні явища, які роблять внесок у зміни стану людини, складно переплетені, мають зворотні зв'язки

та створюють такий відгук електромагнітної активності, який своєю чергою є процесом шифрування вхідною інформацією на різних рівнях діяльності головного мозку. У зв'язку з цим не мають сенсу навіть спроби структурного аналізу енцефалограм з метою виявлення окремих відгуків та їх класифікації мозкової активності та спроби прояву зв'язків між поточною та змінами у навколишньому середовищі та типовому електромагнітному відгуку, який може бути поставлений їм у відповідність[3].

Однак дослідники припускають, що дана ситуація не є безнадійна і є можливість проведення інтелектуального аналізу на основі статистичних матеріалів множин окремих електроенцефалограм та виявлення доведеної залежності між деякими типовими показниками електромагнітної активності та явищами, які їх викликають.

На початку досліджень було поставлено першу гіпотезу. Передбачається, що електроенцефалограма конкретної людини містить інформацію про індивідуальні властивості досліджуваного[4]. Поставлено завдання класифікації електроенцефалограм за ознаками належності контрольного зрізу ритмів мозкової активності конкретної випробуваної особи.

Для цього пропонується порівняльний кластерний аналіз багатовимірних даних пакетів електроенцефалограм для одного і того ж типового стану випробуваного.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Починаючи з 20 століття, складність мозку вивчалася, і в останні роки наука про мозок продовжує залишатися гарячою темою для досліджень [5]. Аналіз електричної активності мозку є однією з головних сфер інтересів науки про мозок [6]. Як результат, електроенцефалограма є важливою для аналізу науки про мозок і часто використовується в різних областях, пов'язаних з мозком [7, 8].

Попередні дослідження показують, що електроенцефалограми можуть містити інформацію про індивідуальні особливості людини [9, 10, 11]. Основною гіпотезою є припущення, що ЕЕГ містить унікальні біометричні дані, які можна використовувати для ідентифікації. Більшість досліджень зосереджені на традиційних підходах аналізу частотних характеристик, проте такі методи не завжди забезпечують необхідну точність [12].

ЕКСПЕРИМЕНТИ

Експерименти проводяться на одному і тому ж електроенцефалографі, який записував інформацію по 22 каналам відведення. при цьому застосовувалася однакова схема монтажу для двох випробовуваних. Для кожного випробовуваного проводиться серія записів у стані спокою сидячи, без рухів та зовнішніх подразників. Серія енцефалограм по 4 секунди проводиться для одного монтажу, після чого схема демонтується і збирається знову. На новій схемі монтажу виконується така ж серія процедур з такою ж кількістю зафіксованих електроенцефалограм.

Таким чином, серія з 10 записів при одному монтажі повторювалася 10 разів, що дало дослідникам 100 незалежних точок вимірювання. Така ж процедура проводилася на іншому випробовуваному. Проведення процедур повторювалося кілька разів і при цьому обов'язково підтримувалися умови ідентичності стану випробовуваних.

З отриманих інформативних даних енцефалограм ми можемо скористатися трьома таблицями: таблицею амплітуд, яка містить дані про електричну активність мозку у часі по відповідним відведенням на схемі монтажу; таблицями потужності з обробленими статистичними даними та таблицею частот яка представлені на рисунку 1.

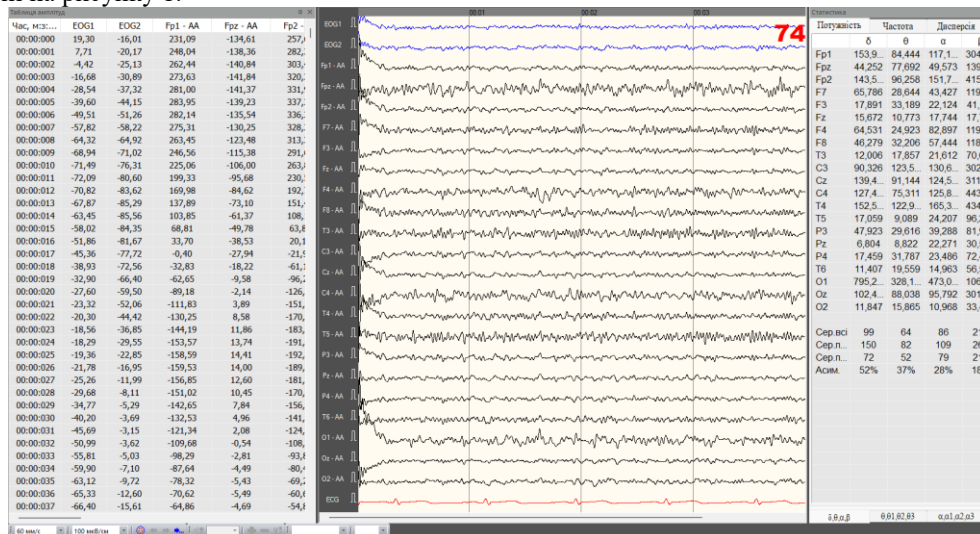


Рис.1. Зовнішній вигляд програмного застосування, з таблицями амплітуд, частот та потужності

Ці дані були експортовані в таблиці Excel Далі зачитані як вхідні дані для статистичного аналізу. На рисунку 2 наведено приклад як проводились розрахунки таблиці потужності та таблиці частот використовуючи алгоритм швидкого перетворення Фур'є [13].

ШПФ-аналіз

Частотні границі:
 Нижня: 1.0 Гц
 Верхня: 30.0 Гц
 Крок: 0.50 Гц

Тривалість епохи: 2000 мсек

Відсоток перекриття: 100

Початкові дані для ШПФ:
☒ фільтрована EEG
☐ нативна EEG

Величина, що аналізується:
☒ Потужність
☐ Амплітуда

Вікно даних:
 Прямокутне
 ПФ (PP)

Дискретно-часова функція:
 Амплітуда
 Відліки

☒ Розрахунок області екрана по видимій області в аналізі

Рис.2. Параметри налаштування алгоритму швидкого перетворення Фур'є

МЕТОД КЛАСТЕРНОГО АНАЛІЗУ

Метод кластеризації базується на розрахунку барицентра кожного каналу відведення. Визначаємо середнє значення амплітуд за всіма відведеннями з таблиці амплітуд і обчислюємо значення барицентра як множину точок по каналах відведення, віддалених на найменшу евклідову відстань від усереднених точок усіх електроенцефалограм кластера. На рисунку 3 представлено обчислювальний метод визначення даних барицентра з 22 відведення для кластера з десяти незалежних енцефалограм.

$$AvAv(nv) := \frac{\sum_{i=0}^{Ng} \left[(AvNCM^{(nv)T})_{0,i} \right]}{Ng + 1}$$

BCAvNC _{nv}	AvAv(nv)
-0.713	-0.806
0.498	0.496
-1.267	-1.43
1.291	1.345
-1.196	-1.239
-0.621	-0.697
-3.043	-3.125
-0.086	-0.181
1.895	2.004
-0.197	-0.25
0.316	0.263
2.215	2.147
-1.285	-1.343
2.039	1.977
-1.373	-1.441
...	...

```

BCAvNC :=
dist0 ← 1000000000
nv1 ← 0
R0 ← 0
while nv1 ≤ 22
    x ← min(AvNCM(nv1)T)
    step ← (max(AvNCM(nv1)T) - min(AvNCM(nv1)T)) / 100
    while x ≤ max(AvNCM(nv1)T)
        ap ← ∑i=0Ng [x - (AvNCM(nv1)T)0,i]
        if ap ≤ dist0
            Rnv1 ← x
            dist0 ← ap
        x ← x + step
    nv1 ← nv1 + 1
    dist0 ← 1000000000
R
    
```

Рис.3. Алгоритм розрахунку барицентра енцефалограм

Кластеризація проводиться для кожного випробуваного шляхом усереднення всіх показників електроенцефалограм для кожного відведення, отриманих у стані спокою. Як інструментальні засоби для проведення статистичного аналізу автори обрали ефективний застосунок для обробки масивів даних. Обробка матриці таблиць у MathCad цілком відповідає вимогам аналізу. Отримана сигнатура барицентра за кожним відведенням наведена на рисунку 4.

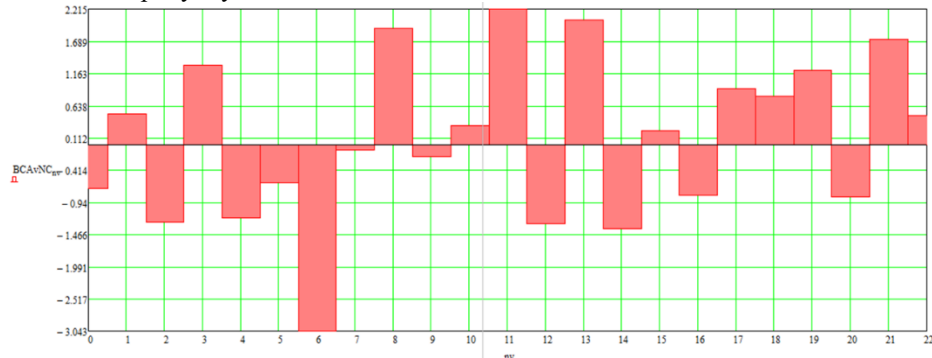


Рис.4. Сигнатура барицентра амплітуд випробуваного Andrei

Далі отримана таблиця барицентра, усереднених статистичних даних за потужністю та частотами ритмів для кожного відведення зберігається у файлах Excel для подальшої обробки.

Після створення кластерів сигнатур амплітуд, потужностей і частот проводиться процедура контролю електроенцефалограм двох випробуваних. На рисунку 5 наведено приклади співвідношень порівняння з кластерними показниками.

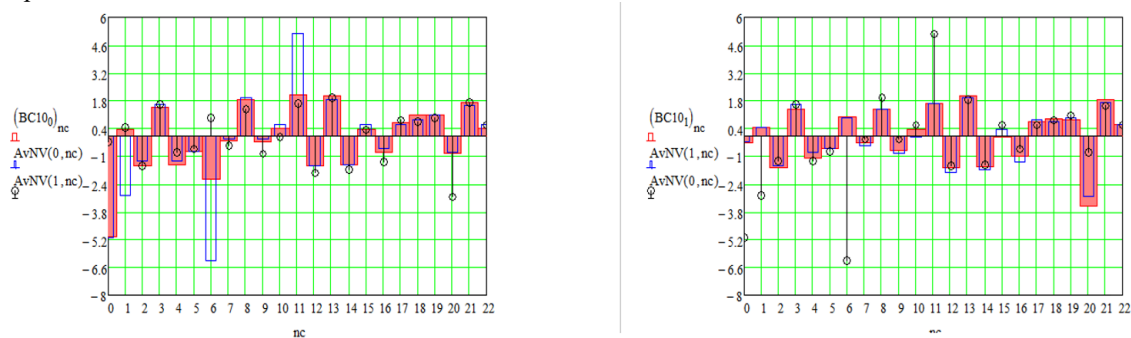


Рис.5. Співвідношень порівняння кластерних показників двох випробуваних

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Результати порівняльних досліджень для випробуваних "Andre" та "Vlad" показані на рисунку 6. Контрольна електроенцефалограма випробуваного Andre відхиляється від його барицентра Andre на 49%, а від барицентра Vlad на 54%. Це підтверджує належність контрольної енцефалограми випробуваному A.

Контрольна енцефалограма Vlad знаходиться на меншій відстані від його барицентра Vlad (77%), ніж від барицентра Andre (218%). Порівняння дистанцій барицентрів також підтверджує належність контрольної енцефалограми до випробувального Vlad. Таким чином, енцефалограми Vlad та Andre за барицентрами класифіковано правильно.

Той самий результат спостерігається і для сигнатур показника потужності SP і частоти SF. Таким чином по показниках трьох значень відбувається підтвердження приналежності електроенцефалограми до особи. Наведений приклад не проходив спеціальної підготовки та був обраний абсолютно випадково із загального масиву енцефалограм. Отримані результати повністю підтверджують можливість ідентифікації особи.

	0	"Andre"	"Vlad"
"% відх BC A"		49.497	218.481
"% відх BC V"		54.891	77.923
"% відх SP A"		1.844	45.026
"% відх SP V"		43.182	7.143×10^{-5}
"% відх SF A"		0.242	1.21
"% відх SF V"		0.55	0.418

Рис.6. Порівняльна таблиця відсоткових відхилень контрольних енцефалограм від барицентрів за амплітудними (BC), потужностями (SP) та частотними характеристиками (SF)

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМКУ

У результаті дослідження було підтверджено гіпотезу про наявність індивідуальних біометричних характеристик у даних електроенцефалограм, які можна використовувати для ідентифікації особистості. Запропонований метод кластерного аналізу показав точність розпізнавання належності контрольної ЕЕГ за трьома статистичними таблицями до конкретного суб'єкта. Результати експериментальних досліджень підтверджують стабільність ідентифікаційних ознак та ефективність обраного підходу. Перевагою методу є його простота, а також можливість реалізації на основі доступних статистичних інструментів. У подальших дослідженнях планується:

- адаптація методів кластерного аналізу до умов змін емоційного або фізіологічного стану суб'єкта;
- впровадження алгоритмів автоматичного розпізнавання та порівняння сигнатур у реальному часі;
- створення ментального інтерфейсу мозок-комп'ютер [14, 15].

References

1. J. Fan, L. Fang, J. Wu, Y. Guo, and Q. Dai, "From brain science to artificial intelligence," *Engineering*, vol. 6, pp. 248–252, 2020.
2. O. Sporns, "Structure and function of complex brain networks," *Dialogues Clin. Neurosci.*, vol. 15, pp. 247–262, 2013.
3. F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, 2018.
4. R. Kanai, B. Bahrami, and G. Rees, "Human parietal cortex structure predicts individual differences in perceptual rivalry," *Current Biology*, vol. 20, no. 18, pp. 1626–1630, 2010.
5. R. D. Lane, L. Ryan, L. Nadel, and L. Greenberg, "Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science," *Behav. Brain Sci.*, vol. 38, p. e1, 2015.
6. F. L. da Silva, "EEG and MEG: Relevance to neuroscience," *Neuron*, vol. 80, pp. 1112–1128, 2013.
7. F. L. Da Silva, "EEG: Origin and measurement," in *EEG-fMRI: Physiological Basis, Technique, and Applications*, Berlin/Heidelberg, Germany: Springer, 2023, pp. 23–48.
8. A. Saibene, M. Caglioni, S. Corchs, and F. Gasparini, "EEG-Based BCIs on motor imagery paradigm using wearable technologies: A systematic review," *Sensors*, vol. 23, p. 2798, 2023.
9. V. H. C. de Albuquerque, R. Damaševičius, J. M. R. S. Tavares, and P. R. Pinheiro, "EEG-Based biometrics: Challenges and applications," *Computational Intelligence and Neuroscience*, 2018, doi: 10.1155/2018/5483921.
10. Hema C.r, Elakkiya.a, Paulraj Mp, "Biometric identification using electroencephalography," *International Journal of Computer Applications*, vol. 106, no. 15, pp. 1–7, 2014, doi: 10.5120/18596-9845. [Online]. Available: <https://www.ijcaonline.org/archives/volume106/number15/18596-9845>
11. V. H. C. de Albuquerque, R. Damaševičius, J. M. R. S. Tavares, and P. R. Pinheiro, "EEG-Based biometrics: Challenges and applications," *Computational Intelligence and Neuroscience*, 2018, doi: 10.1155/2018/5483921.
12. J. C. Monsy and A. P. Vinod, "EEG-based biometric identification using frequency-weighted power feature," *IET Biometrics*, vol. 9, no. 1, pp. 43–51, 2020, doi: 10.1049/iet-bmt.2019.0158.
13. Dian C. R. Novitasari1, Suwanto2, M. Hasan Bisri3, Ahmad Hanif Asyhar, "Classification of EEG signals using fast Fourier transform (FFT) and adaptive neuro-fuzzy inference system (ANFIS)," *Mantik*, vol. 5, no. 1, pp. 35–44, 2019, doi: 10.15642/mantik.2019.5.1.35-44. [Online]. Available: <https://core.ac.uk/download/pdf/235583696.pdf>
14. R. Dunne, T. Morris, and S. Harper, "A survey of ambient intelligence," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–27, 2021.
15. X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, and P. Marttinen, "EEG based emotion recognition: A tutorial and review," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–57, 2015.