

Лифар В.О., Лифар О.К., Захожай О.І.

МЕТОД АСОЦІАТИВНОЇ КЛАСТЕРИЗАЦІЇ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ ІНДИВІДУАЛЬНИХ ОЗНАК РОЗУМОВОЇ АКТИВНОСТІ ЛЮДИНА ЗА ДАНИМИ ЕНЦЕФАЛОГРАМ

Стаття присвячена вирішенню актуальної задачі розробки інформаційної технології визначення індивідуальних ознак розумової активності людини для реалізації в апаратно-програмних інтерфейсів між людиною та комп'ютером з використанням класифікаційних засобів штучного інтелекту та застосування засобів об'єктивного контролю енцефалограм. В статті приводяться стислі дані результатів досліджень на основі 24-канального електроенцефалографа Braintest-24, як засобу об'єктивного контролю розумової діяльності людини. Автори ставили за мету перевірити гіпотезу про унікальність ідентифікаційних можливостей електромагнітної активності мозку, завдяки якій можна з високою долею ймовірності ідентифікувати особу піддослідного і проводити процедуру розпізнавання приналежності знятих енцефалограм конкретній особі. При цьому було розроблено математичний метод розпізнавання, який автори назвали методом асоціативної кластеризації. Проведені дослідження підтвердили високий рівень можливостей ідентифікації особистості піддослідного на основі порівняння статистичних даних досліджень конкретних людей, які аналізувалися та узагальнювалися в кластери та утворювали патерни для вирішення задачі класифікації. Отримані результати експериментів підтверджують сформульовану гіпотезу, а розроблені методи обробки енцефалограм дозволяють використовувати їх в задачах класифікації на основі нейронної мережі. В статті визначений тільки початок досліджень. Майбутні дослідження будуть присвячені аналізу можливостей ідентифікації абстрактного рівня мислення людини та рівня емоціональних проявлень.

Ключові слова: інформаційні системи і технології, енцефалографічний аналіз, штучний інтелект, нейронна мережа, інтерфейс, кластеризація, інтелектуальний аналіз даних.

Вступ. Використання електроенцефалографа як засобу об'єктивного контролю за розумовою активністю людини має широкі перспективи при створенні унікальних інтерфейсів взаємодії людина-комп'ютер. Зокрема при вирішенні завдань класифікації абстрактного мислення людини, її емоційних та фізіологічних станів, можна використовувати засоби штучного інтелекту на нейронних мережах та методи інтелектуального аналізу даних, які застосовуються для інтерпретації даних енцефалограм. При цьому можливе проведення операцій автоматичного машинного навчання зі створенням патернів багатовимірних даних, які б ефективно відповідали класам станів людини.

Перша гіпотеза, яку висувують автори цієї роботи, полягає в тому, що масиви даних електричної активності мозку містять зашифровану інформацію про індивідуальність випробуваного. Передбачається, що інтегральна сукупність даних амплітуди, спектральної потужності та частот енцефалограм може бути оброблена статистичними методами, що дозволили б виділити характерні закономірності, що говорять про належність електроенцефалограми кластеру конкретного індивідуума.

Передбачається також, що більш глибока обробка даних при великому наборі статистичних показників дозволить класифікувати та інтерпретувати психоемоційний стан людини, а також виявити закономірності у зміні станів частотних ритмів енцефалограми при зовнішньому візуальному впливі або осмисленні випробуваних чітко виражених символів. Таким чином з'являється можливість створення інтерфейсу, в якому енцефалограф грає роль посередника між людиною та засобами штучного інтелекту. Приклад даних, що отримувалися багатоканальним енцефалографом та підлягали подальшій обробці, проілюстровано на рисунку 1.

Аналіз питання та постановка завдання. Для підтвердження першої гіпотези було поставлено статистичний експеримент із використанням найпростіших методів контролю та статистичної обробки даних електроенцефалографа для двох піддослідних. Для кожного випробуваного була проведена серія експериментів, зняті енцефалограми групами по 10 записів тривалістю 4 секунди кожна, для однакових монтажів енцефалограм разом по 23 відведенням. Усього таких груп було зареєстровано для п'яти незалежних монтажів. Разом по одному випробуваному було отримано по 50 енцефалограм, які відповідали стану спокою і були калібрувальними. Дослідники намагалися максимально уникати будь-яких артефактів, які могли б вносити додаткові шуми енцефалограми, що реєструються.

Для проведення операції порівняння статистичних даних для кожного пацієнта А та V обиралися по 10 енцефалограм для створення кластерів та по 10 енцефалограм для контрольних порівнянь.

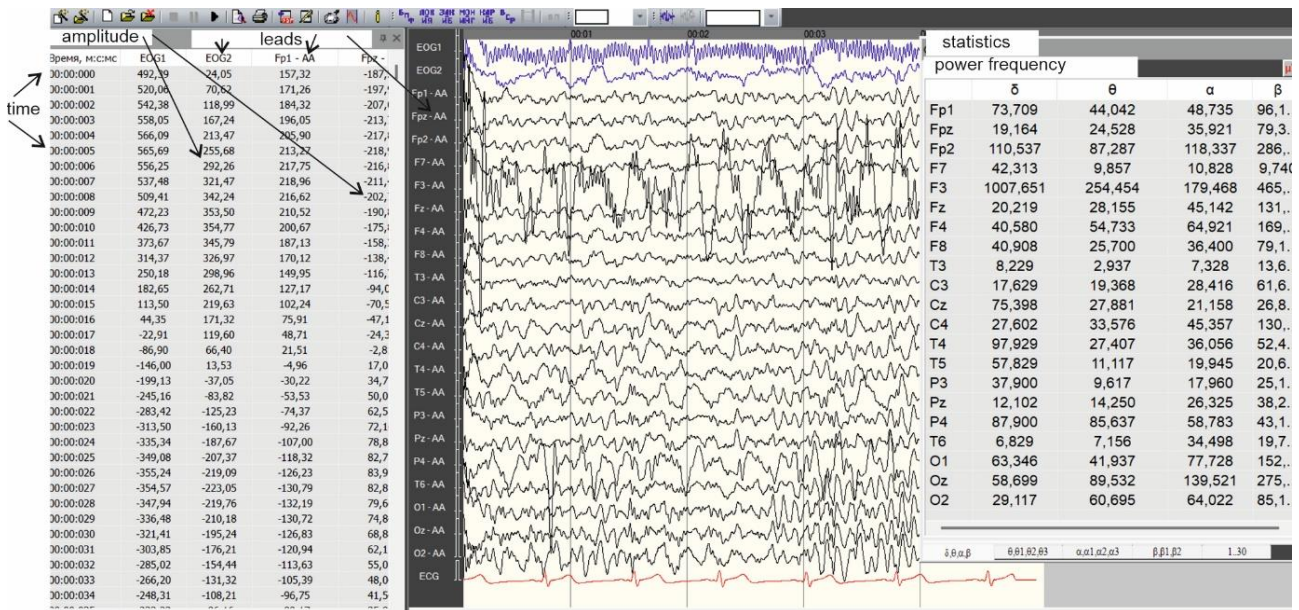


Рисунок 1 – Дані енцефалограми

На рисунку 2 наводиться приклад визначення усереднених даних амплітуди для відведення для 10 енцефалограм пацієнта A і 10 енцефалограм пацієнта V. На рисунку показано кольорове відображення значення амплітуд для обраної енцефалограми та обраного відведення. Зображено дві поверхні відповідних значень для двох пацієнтів. Спостерігається очевидна відмінність, яка надалі обробляється статистичними методами усереднень та порівнянь і може бути інтерпретована як належність певних енцефалограм конкретним пацієнтам. Логічні операції над матрицями дозволяють визначити верхню (OR) та нижню (AND) межі значень амплітуди енцефалограм у кластері з 10 записів. Таким чином, стають відомі дві поверхні, між якими повинні лежати значення енцефалограм досліджуваного пацієнта. При навчанні нейронної мережі амплітуд енцефалограм рекомендується брати їх більшу кількість для одного кластера. Автори взяли 10 енцефалограм виключно для порівняльного прикладу. Кожна енцефалограма усереднена відведенням для 3.000 точок вимірювань (3 секунди). Насправді, якість вимірювань енцефалографа дозволяє оперувати даними по 1 секунді, що значно підвищує швидкість класифікації.

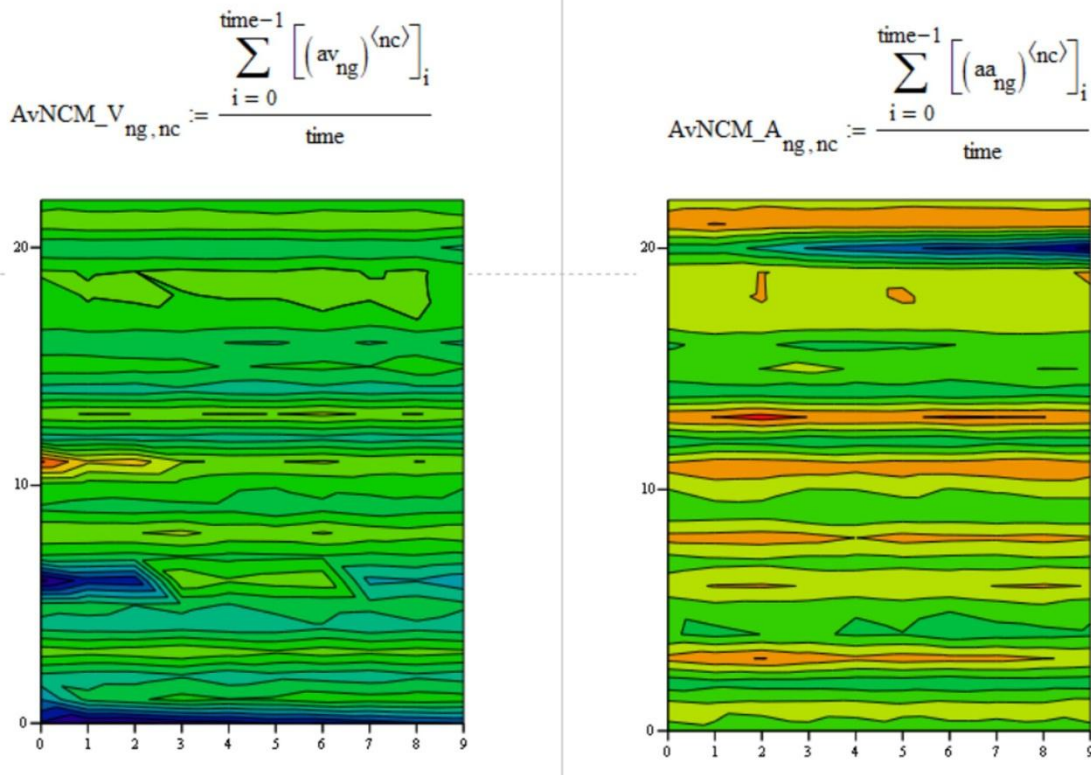


Рисунок 2 - Приклад визначення усереднених даних амплітуди

На рисунку 3 показані поверхні граничних значень амплітуд для двох кластерів.

На рисунку 4 наведено тривимірні значення верхніх та нижніх значень амплітуд кластерного аналізу. Це відображення так званої області значень амплітуд, всередині якої можливий перетин даних контрольних амплітуд енцефалограм пацієнтів. Подальші розрахунки та обробка результатів експерименту показали, що це не призводить до критичних помилок класифікації.

На рисунку 5 наведено розподіл амплітуд за виділеннями для кожного кластера (червоні стовпчики) та значення амплітуд для 10 контрольних енцефалограм кожного кластера (сині відрізки).

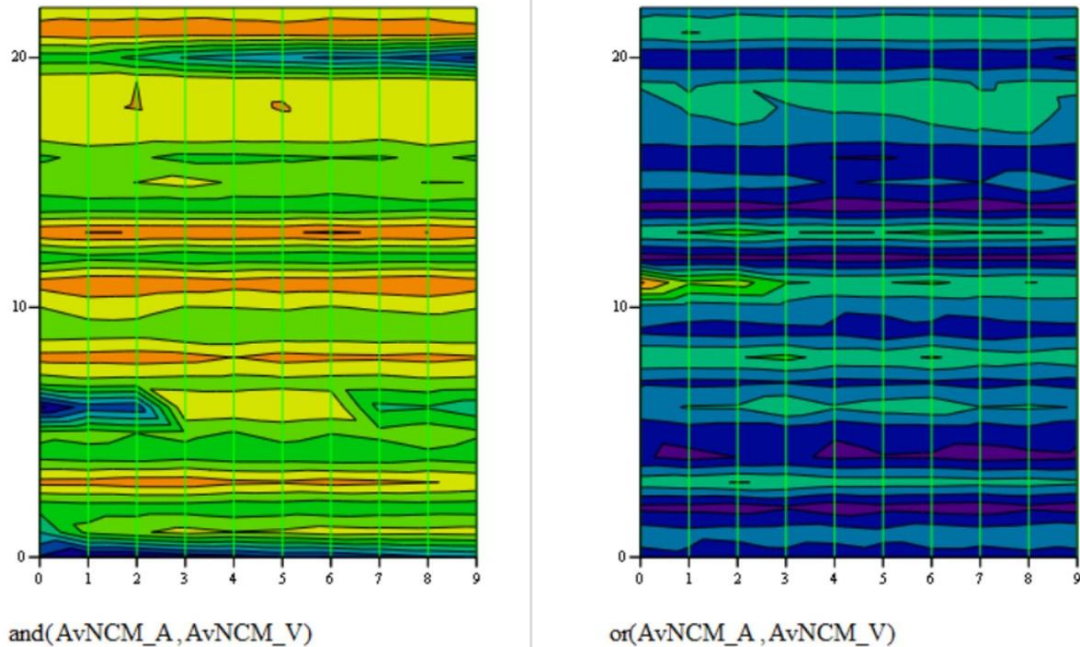


Рисунок 3 – Верхня та нижня область визначення двох кластерів

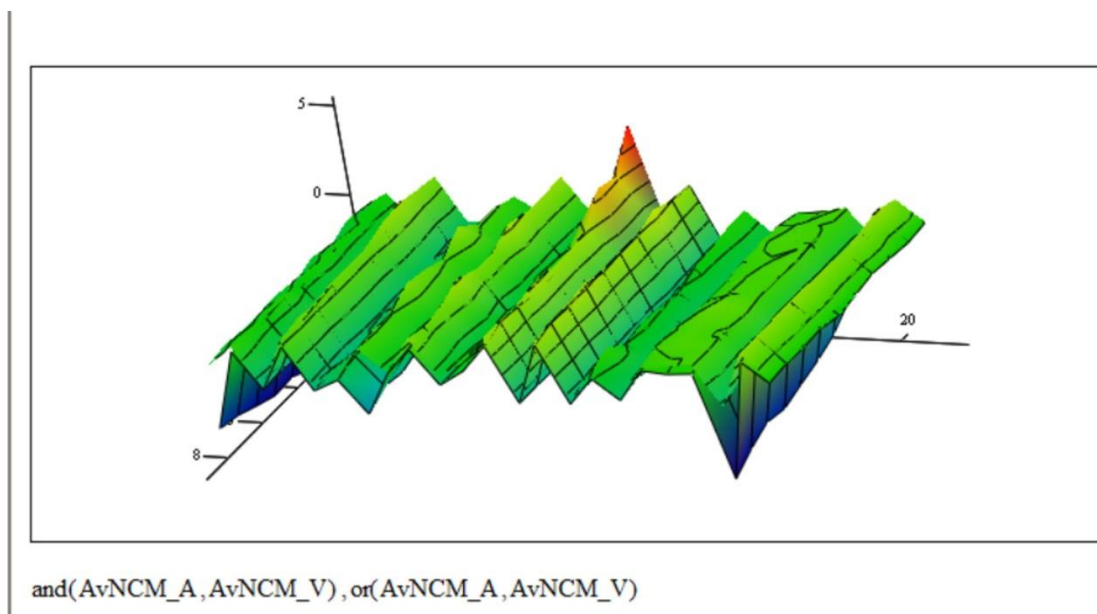


Рисунок 4 – Области визначення значень двох кластерів

Графічне подання показників кластерів на рисунку 5 дозволяє виділити структури кластерів, їх схожість та різницю по каналах відведення. Ця різниця містить визначальну інформацію для того одного й того самого стану пацієнта при знятті енцефалограм в різних експериментах. Наприклад, очевидна різниця відведення 6. Інші відведення, такі як 16, 17, 18, 19 показують близькість та схожість структури діаграм амплітуд. Сині стовпчики на графіках показують накладення середніх амплітуд енцефалограм розподілені за відведенням та узагальнені у кластери.

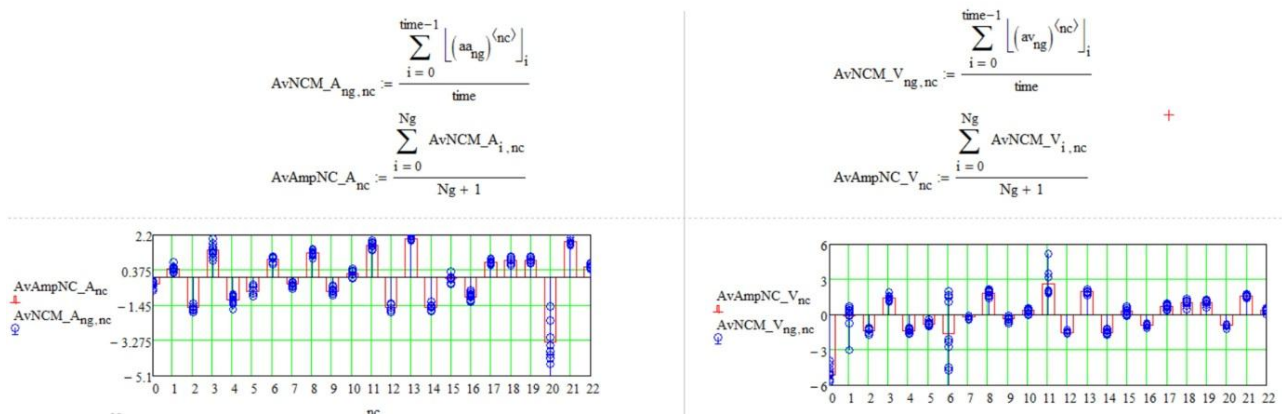


Рисунок 5 – Графіки значень амплітуд щодо відведень для двох кластерів та їх енцефалограм

Далі проводиться порівняння усереднених показників контрольних енцефалограм із показниками кластерів. При цьому для кожного відведення обчислюється дистанція Ейлера. Обчислені показники інтегруються та порівнюються між собою.

На рисунку 6 наведено таблицю порівнянь інтегральних ейлерових дистанцій для 20 контрольних енцефалограм двох пацієнтів А та V. Колонка 1 “Dist A-kA” містить значення ейлерових дистанцій між контрольними енцефалограмами пацієнта А та кластером А, отриманим на підставі статистичних енцефалограм цього пацієнта. Колонка 2 містить розрахункові показники ейлерових дистанцій між контрольними енцефалограмами пацієнта А та кластером, отриманим із статистичних енцефалограм пацієнта V. Очевидно, що дистанція енцефалограм пацієнта А від кластера kV перевищує дистанцію від енцефалограм пацієнта А до кластера kA. Це ж спостерігається і для контрольних енцефалограм пацієнта V.

	0	1	2	3	4
0	0	"Dist A-kA"	"Dist A-kV"	"Dist V-kV"	"Dist V-kA"
1	1	5.22	14.14	12.81	26.24
2	2	5.38	12.61	6.93	20.37
3	3	5.07	13.02	6.94	20.31
4	4	3.03	15.37	8.27	14.7
5	5	3.02	16.16	6.2	11.25
6	6	2.34	15.07	6.75	11.36
7	7	3.98	17.38	7.63	13.14
8	8	2.46	15.51	5.25	12.94
9	9	4.08	16.02	5.54	13.93
10	10	4.37	17.32	7.27	13.92

Рисунок 6 – Порівняння ейлерової дистанції контрольних енцефалограм пацієнтів та кластерів

Усі контрольні виміри проводилися абсолютно у випадковому порядку не містять жодних систематичних відхилень і підтверджують гіпотезу у тому, що енцефалограма містить виражену інформацію про належність її конкретному індивідууму.

Висновки. В результаті проведеного дослідження були отримані наступні основні результати.

Отримані експериментальні дані та методи порівняльного кластерного аналізу підтверджують гіпотезу про те, що серед загального статистичного потоку інформації, представленого в енцефалограмах, можна виділити інформацію про належність контрольної енцефалограми конкретному індивідууму. Це можливо при отриманні статистичних даних та створенні кластерів, характерних для конкретних випробуваних. Експеримент підтверджує стійку систематичну різницю в ейлерових дистанціях між контрольними енцефалограмами та даними кластерів, що дає впевненість отримання подібної інформації та для більшої кількості піддослідних. Ці знання мають значення для розробки інформаційної технології ідентифікації особистостей. Подібні методи можуть бути використані і для завдань класифікації розумових процесів людини на рівні абстрактних уявних об'єктів та психофізіологічних станів людини.

Література

1. Chen, Y.; Wang, W.; Yan, S.; Wang, Y.; Zheng, X.; Lv, C. Application of Electroencephalography Sensors and Artificial Intelligence in Automated Language Teaching. *Sensors* 2024, 24, 6969. <https://doi.org/10.3390/s24216969>
2. Leung, C.; Scarino, A. Reconceptualizing the nature of goals and outcomes in language/s education. *Mod. Lang. J.* 2016, 100, 81–95. [Google Scholar] [CrossRef]
3. Zhang, Y.; Wa, S.; Liu, Y.; Zhou, X.; Sun, P.; Ma, Q. High-accuracy detection of maize leaf diseases CNN based on multi-pathway activation function module. *Remote Sens.* 2021, 13, 4218. [Google Scholar] [CrossRef]
4. Sharma, V.K.; Holbah, W.A. Online Language Assessment the Exception, Not the Rule: For Inclusive Language Learning. *Arab World Engl. J.* 2022. [Google Scholar]
5. Li, Q.; Zhang, Y.; Ren, J.; Li, Q.; Zhang, Y. You Can Use But Cannot Recognize: Preserving Visual Privacy in Deep Neural Networks. *arXiv* 2024, arXiv:2404.04098. [Google Scholar]

References

1. Chen, Y.; Wang, W.; Yan, S.; Wang, Y.; Zheng, X.; Lv, C. Application of Electroencephalography Sensors and Artificial Intelligence in Automated Language Teaching. *Sensors* 2024, 24, 6969. <https://doi.org/10.3390/s24216969>
2. Leung, C.; Scarino, A. Reconceptualizing the nature of goals and outcomes in language/s education. *Mod. Lang. J.* 2016, 100, 81–95. [Google Scholar] [CrossRef]
3. Zhang, Y.; Wa, S.; Liu, Y.; Zhou, X.; Sun, P.; Ma, Q. High-accuracy detection of maize leaf diseases CNN based on multi-pathway activation function module. *Remote Sens.* 2021, 13, 4218. [Google Scholar] [CrossRef]
4. Sharma, V.K.; Holbah, W.A. Online Language Assessment the Exception, Not the Rule: For Inclusive Language Learning. *Arab World Engl. J.* 2022. [Google Scholar]
5. Li, Q.; Zhang, Y.; Ren, J.; Li, Q.; Zhang, Y. You Can Use But Cannot Recognize: Preserving Visual Privacy in Deep Neural Networks. *arXiv* 2024, arXiv:2404.04098. [Google Scholar]

The article is devoted to solving the urgent problem of developing a software-hardware interface between a person and a computer using classification tools of artificial intelligence for interpreting human mental activity based on the use of means of objective control of encephalograms. The article provides concise data on the results of research into the capabilities of a 24-channel, professional electroencephalograph as a means of objective control of human mental activity. The authors of the development aimed to test the hypothesis of the uniqueness of the identification capabilities of the electromagnetic activity of the brain, thanks to which it is possible to identify the subject's personality with a high degree of probability and carry out the procedure for recognizing the belonging of the recorded encephalograms to a specific person. In this case, it was necessary to develop a method of mathematical recognition, which the authors called the method of associative clustering. The conducted studies confirmed the high level of capabilities for identifying the subject's personality based on a comparison of statistical data from studies of specific people, which were analyzed and generalized into clusters and created patterns for solving the classification problem. The obtained experimental results confirm the hypothesis and the developed methods of processing encephalograms allow their use in classification tasks based on a neural network. The article only defines the beginning of the research. Future research will be devoted to the analysis of the possibilities of identifying the abstract level of human thinking and the level of emotional manifestations.

Keywords: information systems and technologies, encephalograph, artificial intelligence, neural network, interface, clustering, intelligent data analysis.

Лифар В.О. - професор кафедри інформаційних технологій та програмування, Східноукраїнський національний університет імені Володимира Даля,

Лифар О.К. - старший викладач кафедри інформаційних технологій та програмування, Східноукраїнський національний університет імені Володимира Даля,

Захожай О.І. - завідувач кафедри інформаційних технологій та програмування, Східноукраїнський національний університет імені Володимира Даля,