

УДК 519.7:616.8

О. К. БІЛОШИЦЬКА

КЛАСИФІКАЦІЯ ПАТОЛОГІЧНИХ ЕЕГ-СИГНАЛІВ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Робота присвячена застосуванню методів машинного навчання при створенні моделі класифікації патологічних сигналів ЕЕГ. Сформовано ефективний набір ознак, які є пріоритетними під час класифікації сигналів з епілептиформною та нормальною активністю головного мозку. Встановлено, що застосування методу опорних векторів дає точність класифікації 80 %, методу лінійного дискримінантного аналізу – 91 %, методу випадкового лісу – 96 % та методу Extra trees – 91 %. Побудована комплексна модель на основі даних методів дала результуючу точність 89 %.

Ключові слова: ЕЕГ, машинне навчання, метод опорних векторів, лінійний дискримінантний аналіз, випадковий ліс, Extra trees.

Работа посвящена применению методов машинного обучения при создании модели классификации патологических сигналов ЭЭГ. Сформирован эффективный набор признаков, которые являются приоритетными во время классификации сигналов с эпилептиформной и нормальной активностью головного мозга. Установлено, что применение метода опорных векторов дает точность классификации 80 %, методу линейного дискриминантного анализа – 91 %, методу случайного леса – 96% и методу Extra trees – 91 %. Построенная комплексная модель на этих методах дала результирующую точность 89 %.

Ключевые слова: ЭЭГ, машинное обучение, метод опорных векторов, линейный дискриминантный анализ, случайный лес, Extra trees.

Epilepsy is the fourth most common neurological problem in the world. When diagnosing epilepsy, the most informative is the registration of EEG, which helps distinguish epileptic seizures from non-epileptic seizures and classify them.

As a result of the preliminary processing of raw data of EEG signals, segments of signals with epileptic activity were isolated separately, and signals without pathological rhythms were divided into segments of 30 seconds duration. A database of 194 signals with epileptiform activity and 194 signals of normal brain activity was obtained.

Performed preprocessing of EEG, an effective set of characteristics has been generated for the classification of signals with epileptiform and normal activity of brain; a model for the classification of pathological EEG signals was created and based on methods of machine learning.

An effective set of characteristics has been generated for the classification of signals with epileptiform and normal activity of brain.

A model for classifying pathological EEG signals was created in the software environment Python 2.7.5. Model consisted of methods Support Vector Machine, Linear Discriminant Analysis, Random Forest and Extra trees. The constructed complex model on these methods has shown the resulting accuracy of 89 %.

Keywords: EEG, machine learning, Support Vector Machine, Linear Discriminant Analysis, Random Forest and Extra trees.

Вступ. Епілепсія є четвертим найбільш поширеним неврологічним захворюванням людини в світі. Кожного року вперше епілептичні напади реєструють близько в 1000 людей. Такі напади істотно обмежують соціальну активність та можливість людей. Проте своєчасне діагностування, детектування та попередження нападів можуть значно полегшити перебіг захворювання та, навіть, врятувати життя [1]. Так, наприклад, знаючи, що за декілька хвилин або годин може відбутись епілептичний напад, людина може зупинити машину, відпроситись з роботи і прийняти ліки, щоб полегшити симптоми.

При діагностуванні епілепсії найбільш інформативною є реєстрація ЕЕГ, яка допомагає відрізнити епілептичні напади від неепілептичних та класифікувати їх.

Аналіз літературних даних та постановка проблеми. Класифікація електроенцефалограм є одним з важливих процесів в біомедичній інженерії. Втома водія – одна з основних причин ДТП у всьому світі, і вважаються серйозною проблемою в останні десятиліття [2]. Hybrid Deep Generic Model (DGM) реалізована в [3] на основі методу опорних векторів SVM (Support Vector Machine) і застосовується для точного виявлення втоми водія.

У запропонованій системі, DGM використовується для вилучення ознак, в той час як SVM – для їх класифікації. Експериментальні результати показали, що запропонована система моніторингу втоми водія забезпечує точність тестування на 73,29 %.

У роботі [4] автори порівнюють ефективність класифікації ЕЕГ-сигналів в системі ВСІ (Інтерфейс «мозок-комп'ютер» (Brain-Computer Interface)) методами ELM (Екстремальне навчання (Extreme Learning Machine)), LDA і SVM. Експериментальні результати показали, що більш високу точність класифікації демонструє метод ELM.

Для того, щоб вирішити проблему нестабільних результатів прогнозування, коли один метод ELM трактується як класифікатор, в цій роботі [5] автори пропонують метод класифікації з використанням ансамблю ELM та LDA (Лінійний дискримінантний аналіз – Linear Discriminant Analysis). Основна ідея, це застосування LDA на кожній підмножині навчальних наборів даних. Таким чином, можна отримати вибірку з великими відмінностями, що збільшує різноманітність даних і зменшує можливі помилки ансамблю.

Дослідження [6, 7] спрямоване на виявлення взаємозв'язку між ЕЕГ-сигналами і емоційним станом людини. В якості стимулів, щоб викликати позитивні, нейтральні та негативні емоції суб'єктів, використовувались відеоролики.

У дослідженні [8] автори пропонують систему (рис. 1) для ідентифікації людини з використанням ЕЕГ-сигналів, отриманих від таких недорогих пристроїв. Сигнал ЕЕГ спочатку проходить попередню обробку для видалення шумів і артефактів за допомогою смугового фільтра. Потім ці сигнали розділяються на окремі відрізки.

© О. К. Білошицька. 2017

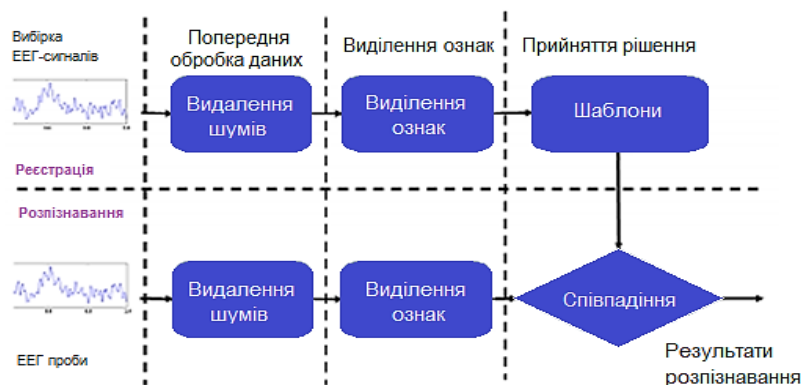


Рис. 1 – Блок-схема системи розпізнавання EEG-сигналів

Ціль та задачі дослідження. Епілепсія може діагностуватись в будь-якому віці. Ідентифікація можлива шляхом дослідження сигналів електроенцефалограми. У роботі [8] автори пропонують метод обробки EEG-сигналів для розпізнавання епілептичних нападів, який складається з вейвлет-аналізу та системи машинного навчання. Використовувалася значна кількість класифікаторів, але найбільшу точність показав LDA-класифікатор (приблизно 93 %), а також метод опорних векторів SVM (70 %).

В дослідженні [9] автори пропонують контрольований метод машинного навчання для класифікації EEG-сигналів нормальної активності мозку, перед епілептичним нападом і під час нападу (рис. 2).

Виходячи з вищезазначеного, було сформульовано мету нашого дослідження, яка полягає в створенні моделі класифікації патологічних EEG-сигналів на основі методів машинного навчання.

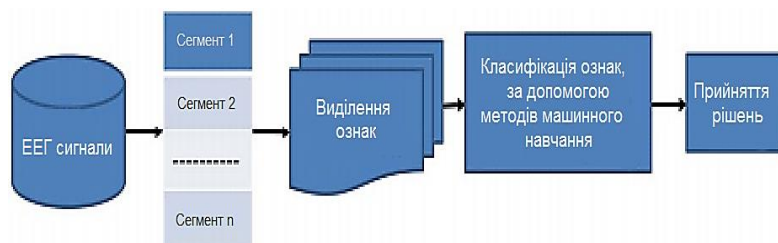


Рис. 2 – Блок-схема системи класифікації епілепсії

Матеріали та методи дослідження патологічних сигналів EEG. База EEG-сигналів з епілептичними нападами була отримана від 24-х пацієнтів: 5 чоловіків, віком від 22 до 36 років, та 19 жінок, віком від 20 до 40 років. Тривалість реєстрації сигналу складає 1 годину. EEG-сигнали були отримані від 16 каналів, із міжнародною системою накладання електродів 10-20 і оцифровані з частотою 256 значень за секунду. Також була надана інформація, для кожного пацієнта, про кількість епілептичних нападів в одному сигналі, час їх початку і закінчення.

На етапі попередньої обробки даних були виділені відрізки сигналу, в яких зареєстрована епілептиформна активність. Тривалість нападів варіюється від десятків секунд до декількох хвилин [10]. Були видалені всі наявні пропуски в сигналі та

артефакти. Сигнали без патологічної активності були поділені на дві групи: відрізки тривалістю 30 та 60 секунд.

В результаті було отримано базу даних з 194 сигналів з епілептиформною активністю та 194 сигналів нормальної активності мозку, з рівною кількістю каналів для кожного сигналу

Використовуючи «плаваюче вікно» тривалістю 10 секунд, сигнал попередньо поділявся на епохи.

Набір ознак розраховувався для кожної епохи:

- спектральна та ентропія Шеннона на шести частотних смугах (дельта, тета, альфа, бета та гама-ритм);
- частота спектрального краю;
- ентропія Шеннона для діадних частотних смуг;
- матриця кореляції між частотними смугами та її власні значення;
- матриця кореляції між каналами EEG-сигналу та її власні значення;
- кореляція між каналами;
- фрактальна розмірність [10];
- параметри Hjorth (активність, мобільність, і складність);
- статистичні характеристики (асиметрія, коефіцієнт ексцесу, середні значення для кожного каналу);
- перетин нуля (Zero Crossing).

Для дослідження були обрані методи машинного навчання, що безпосередньо вирішують задачу класифікації, а саме метод опорних векторів, лінійний дискримінантний аналіз, випадковий ліс та Extra trees.

Метод опорних векторів заснований на ідеї поділу простору об'єктів, на підпростори, відповідні класам. У разі бінарної класифікації навчання методу зводиться до пошуку гіперплощини з деякою товщиною, яка є математичною сутністю методу, що розділяє об'єкти різних класів навчальної вибірки.

Лінійний дискримінантний аналіз – це метод пошуку лінійної комбінації змінних, яка найкращим чином розділяє два або більше класів.

Основна ідея алгоритму випадкового лісу полягає в використанні ансамблю дерев прийняття рішень. Дерева в ансамблі будуються один від одного незалежно. Фінальна класифікація проводиться за допомогою «голосування», тобто підсумковим класом оголошується той клас, за який проголосувало найбільша кількість дерев.

В екстремально випадкових деревах Extra trees при поділі вузла вибирається випадкова підмножина ознак пороги вибираються випадково для всіх ознак, і вже серед них вибирається найкращий.

Результати дослідження методів машинного навчання під час вирішення задачі класифікації. Задача класифікації полягає в тому, що є безліч

об'єктів, в даному випадку сигналів, розділених деяким чином на класи (епілептична та нормальна активність) – двокласова класифікація. Визначена скінченна множина об'єктів, для яких відомо, до яких класів вони належать, - навчальна вибірка. Класова приналежність інших об'єктів не відома – тестова вибірка. Потрібно побудувати алгоритм, здатний класифікувати довільний об'єкт з тестової вибірки. У машинному навчанні завдання класифікації відноситься до розділу навчання з учителем. На рис. 3 зображена блок-схема такої системи класифікації.

Всі дослідження були проведені за допомогою програмного забезпечення Python 2.7.5, з використанням бібліотек.

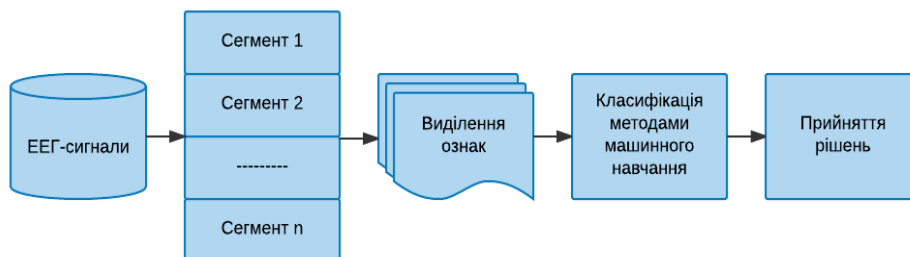


Рис. 3 – Блок-схема системи класифікації епілептичних нападів

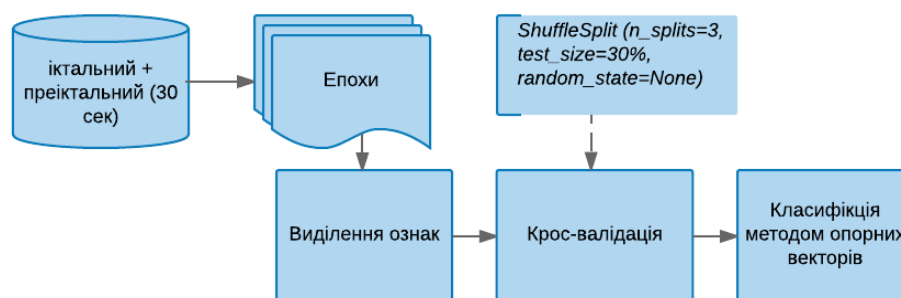


Рис. 4 – Блок-схема поетапного виконання моделювання

Обговорення результатів моделі класифікації патологічних ЕЕГ-сигналів. Під час подальшого підбору методів машинного навчання для поділу даних на навчальну та тестову вибірку використовувалась функція крос-валідації `train_test_split`. Параметр ядра класифікатора був змінений з лінійного на поліноміальний (`poly`). Точність класифікації зростає до 78,23 %.

Для підвищення якості роботи алгоритмів, був проведений відбір оптимальних характеристик сигналу, поступово видаляючи набір ознак, по одному за раз. В результаті чого, були видалені такі ознаки: фрактальна розмірність, ентропія Шеннона на шести частотних смугах та для діадних частотних смуг і матриця кореляції між каналами. Після чого кількість ознак зменшилась від 232 до 186. Точність класифікації складала для методу опорних векторів – 78 %, лінійного дискримінантного аналізу – 88 %, випадкового лісу – 89 % та Extra trees – 84 %. Показники покращились для всіх класифікаторів, окрім методу опорних векторів, значення точності для якого залишились незмінними.

На цьому ж наборі ознак була протестована вибірка сигналів з нормальною активністю головного мозку тривалістю 60 секунд. Послідовність аналізу

Перший результат класифікації ЕЕГ-сигналів був отриманий на наборі даних, який складався з розрахованих ознак для епілептичних нападів та сигналів нормальної активності мозку тривалістю 30 секунд.

Перед тим як подавати набір даних на вхід алгоритму для класифікації, необхідно розділити загальну вибірку на навчальну, на якій безпосередньо здійснюватиметься навчання, та тестову, об'єкти якої, алгоритм самостійно розподіляє між класами.

Такий поділ вибірки виконується функціями крос-валідації. Як класифікатор використовувався метод опорних векторів з лінійним ядром та параметром `C` за замочуванням (рис. 4).

сигналів не змінилась, поділ вибірки так само здійснювався за допомогою функції `train_test_split`. Точність складала для методу опорних векторів – 79 %, лінійного дискримінантного аналізу – 86 %, випадкового лісу – 71 % та Extra trees – 66 %. Отримані значення значно знизились, що показує кращу ефективність у використанні сигналів без патологічної активності головного мозку тривалістю в 30 секунд.

Після здійсненого відбору ознак, набір характеристик має такий склад:

- частота спектрального краю;
- матриця кореляції між частотними смугами;
- кореляція між каналами ЕЕГ-сигналу;
- параметри Hjorth (активність, мобільність, і складність);
- статистичні характеристики (асиметрія, коефіцієнт ексцесу, середнє значення для кожного каналу);
- перетин нуля.

Однією із можливостей алгоритмів класифікації є оцінювання важливості ознак. Тож для зменшення кількості ознак та виділення найінформативніших, була використана функція `SelectFromModel`, яка трансформує вибірку на основі важливості ознак, оцінених методом Extra trees. З її допомогою було

відфільтровано 47 ознак з найкращим показником вагового коефіцієнта. Чим вище ваговий коефіцієнт, тим важливіша ознака. В результаті відбору найважливішою ознакою для класифікації сигналів з епілептиформною активністю стала частота спектрального краю, з іншого боку, найгірші показники мають ознаки асиметрії та середнє значення, які були видалені.

В результаті оцінювання важливості ознак, точність класифікації підвищилась для всіх методів. Цей результат свідчить про те, що «якість» та «унікальність» даних дуже важливі для їх ефективного аналізу.

На основі отриманих результатів, була реалізована модель для автоматичної класифікації ЕЕГ-сигналів. На рис. 5 зображена структура даної системи класифікації.

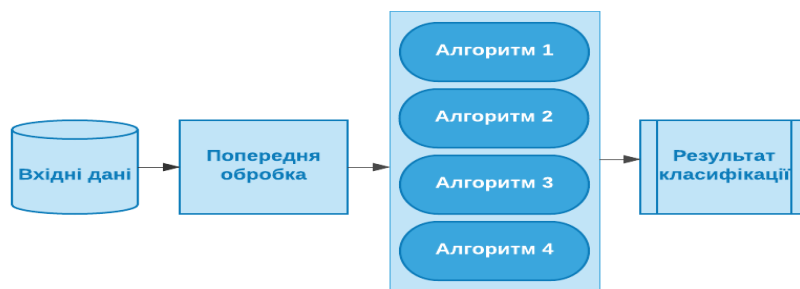


Рис. 5 – Блок-схема результуючої моделі класифікації

Фінальна точність класифікації склала 89% і була розрахована за формулою:

$$P = 1/4 \cdot (\text{Алгоритм 1} + \text{Алгоритм 2} + \text{Алгоритм 3} + \text{Алгоритм 4})$$

де Алгоритм 1 – точність класифікації за допомогою методу SVM; Алгоритм 2 – точність класифікації за допомогою LDA; Алгоритм 3 – точність класифікації за допомогою методу RF; Алгоритм 4 – точність класифікації за допомогою методу Extra trees.

Висновки. За допомогою оцінювання важливості ознак функцією SelectFromModel, був відібраний найефективніший набір ознак для детектування епілептиформної активності, а саме: частота спектрального краю, кореляція між каналами, матриця кореляції між частотними смугами, параметри Hjorth (активність, мобільність, і складність), коефіцієнт ексцесу, перетин нуля.

Класифікація була здійснена методом опорних векторів, лінійного дискримінантного аналізу, випадкового лісу та Extra trees та отримані такі показники точності: 80 %, 91 %, 96 %, 91 % відповідно.

На основі проведених експериментів була побудована модель для автоматичної класифікації ЕЕГ-сигналів за допомогою ансамблю таких ме-

тодів машинного навчання. Результуюча точність склала 89 %.

Список літератури:

1. Білошицька, О. К. Нелінійна динаміка як інструмент прогнозування патологічних змін на електроенцефалограмі [Текст] / О. К. Білошицька // Вісник НТУ «ХПІ». Серія: Механіко-технологічні системи та комплекси. – 2016. – № 50 (1222). – С. 79–83.
2. Доля, В. К. Вплив інформаційного навантаження на параметри основної діяльності водіїв (збуджувальні процеси) [Текст] / В. К. Доля, І. П. Енглезі, І. А. Афанасєва // Східно-Європейський журнал передових технологій. – 2011. – Т. 1, № 2 (49). – С. 65–68. – Режим доступу: <http://journals.urau.ua/eejet/article/view/1887/1782>
3. San, P. P. EEG-based driver fatigue detection using hybrid deep generic model [Text] / P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig, H. Nguyen // 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). – 2016. doi: [10.1109/embc.2016.7590822](https://doi.org/10.1109/embc.2016.7590822)
4. Tan, P. Applying Extreme Learning Machine to classification of EEG BCI [Text] / P. Tan, W. Sa, L. Yu // 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). – 2016. doi: [10.1109/cyber.2016.7574827](https://doi.org/10.1109/cyber.2016.7574827)
5. Han, M. EEG Signals Classification Based on Wavelet Packet and Ensemble Extreme Learning Machine [Text] / M. Han, Z. Sun, J. Wang // 2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI). – 2015. doi: [10.1109/mcsi.2015.30](https://doi.org/10.1109/mcsi.2015.30)
6. Liu, Q. Multi-scale wavelet kernel extreme learning machine for EEG feature classification [Text] / Q. Liu, X. Zhao, Z. Hou, H. Liu // 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). – 2015. doi: [10.1109/cyber.2015.7288175](https://doi.org/10.1109/cyber.2015.7288175)
7. Zhu, J.-Y. EEG-based emotion recognition using discriminative graph regularized extreme learning machine [Text] / J.-Y. Zhu, W.-L. Zheng, Y. Peng, R.-N. Duan, B.-L. Lu // 2014 International Joint Conference on Neural Networks (IJCNN). – 2014. doi: [10.1109/ijcnn.2014.6889618](https://doi.org/10.1109/ijcnn.2014.6889618)
8. Dilber, D. EEG based detection of epilepsy by a mixed design approach [Text] / D. Dilber, J. Kaur // 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). – 2016. doi: [10.1109/rteict.2016.7808067](https://doi.org/10.1109/rteict.2016.7808067)
9. Bashar, M. K. Epileptic seizure classification from intracranial EEG signals: A comparative study EEG-based seizure classification [Text] / M. K. Bashar, F. Reza, Z. Idris, H. Yoshida // 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). – 2016. doi: [10.1109/iecbes.2016.7843422](https://doi.org/10.1109/iecbes.2016.7843422)
10. Білошицька, О. К. Дослідження показників епілептичних ЕЕГ-сигналів за допомогою методів нелінійної динаміки [Текст] / О. К. Білошицька, Т. А. Клименко // Вісник НТУ «ХПІ». Серія: Механіко-технологічні системи та комплекси. – 2017. – № 19 (1241). – С. 30–34.

Bibliography (transliterated):

1. Biloshytska, O. K. (2016). Nelineynna dynamika yak instrument prognuzuvannya patolohichnykh zmin na elektroentsefalogrami. Visnyk NTU «KhPI». Seriya: Mekhaniko-tekhnologichni systemy ta komplekxy, 50 (1222), 79–83.
2. Dolia, V. K., Enhlezi, I. P., Afanasieva, I. A. (2009). Influence of information load on the basic parameters of drivers activity (exciting process). Eastern-European Journal of Enterprise Technologies, 1 (2 (49)), 65–68. Available at: <http://journals.urau.ua/eejet/article/view/1887/1782>
3. San, P. P., Ling, S. H., Chai, R., Tran, Y., Craig, A., Nguyen, H. (2016). EEG-based driver fatigue detection using hybrid deep generic model. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi: [10.1109/embc.2016.7590822](https://doi.org/10.1109/embc.2016.7590822)

4. Tan, P., Sa, W., Yu, L. (2016). Applying Extreme Learning Machine to classification of EEG BCI. 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). doi: [10.1109/cyber.2016.7574827](https://doi.org/10.1109/cyber.2016.7574827).
5. Han, M., Sun, Z., Wang, J. (2015). EEG Signals Classification Based on Wavelet Packet and Ensemble Extreme Learning Machine. 2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI). doi: [10.1109/mcsi.2015.30](https://doi.org/10.1109/mcsi.2015.30).
6. Liu, Q., Zhao, X., Hou, Z., Liu, H. (2015). Multi-scale wavelet kernel extreme learning machine for EEG feature classification. 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). doi: [10.1109/cyber.2015.7288175](https://doi.org/10.1109/cyber.2015.7288175).
7. Zhu, J.-Y., Zheng, W.-L., Peng, Y., Duan, R.-N., Lu, B.-L. (2014). EEG-based emotion recognition using discriminative graph regularized extreme learning machine. 2014 International Joint Conference on Neural Networks (IJCNN). doi: [10.1109/ijcnn.2014.6889618](https://doi.org/10.1109/ijcnn.2014.6889618).
8. Dilber, D., Kaur, J. (2016). EEG based detection of epilepsy by a mixed design approach. 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). doi: [10.1109/rteict.2016.7808067](https://doi.org/10.1109/rteict.2016.7808067).
9. Bashar, M. K., Reza, F., Idris, Z., Yoshida, H. (2016). Epileptic seizure classification from intracranial EEG signals: A comparative study EEG-based seizure classification. 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). doi: [10.1109/iecbes.2016.7843422](https://doi.org/10.1109/iecbes.2016.7843422).
10. Biloshytska, O. K., Klymenko, T. A. (2017). Doslidzhennia pokaznykh epileptychnykh EEH-sygnaliv za dopomohoiu metodiv nelineinnoi dynamiky. Visnyk NTU «KhPI». Seriya: Mekhaniko-tehnologichni systemy ta komplekсы, 19 (1241), 30–34.

Надійшла (received) 14.12.2017

Бібліографічні описи / Библиографические описания / Bibliographic descriptions

Класифікація патологічних EEG-сигналів за допомогою методів машинного навчання/ Білошицька О. К. //Bulletin of NTU “KhPI”. Series: Mechanical-technological systems and complexes. – Kharkov: NTU “KhPI”, 2017. – № 44 (1266).– P.35–39. – Bibliogr.:10. – ISSN 2079-5459

Классификация патологических ЭЭГ-сигналов с помощью методов машинного обучения/ Белошицкая О. К. //Bulletin of NTU “KhPI”. Series: Mechanical-technological systems and complexes. – Kharkov: NTU “KhPI”, 2017. – № 44 (1266).– P.35–39. – Bibliogr.:10. – ISSN 2079-5459

Classification of pathological EEG signals using machine learning methods/ Biloshytska O. //Bulletin of NTU “KhPI”. Series: Mechanical-technological systems and complexes. – Kharkov: NTU “KhPI”, 2017. – № 44 (1266).– P.35–39. – Bibliogr.:10. – ISSN 2079-5459

Відомості про авторів / Сведения об авторах / About the Authors

Білошицька Оксана Костянтинівна – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Факультет біомедичної інженерії, старший викладач кафедри біомедичної інженерії; м. Київ, проспект Перемоги, 37, тел.: 099-606-19-75, 044-204-96-84; e-mail: o.biloshytska@kpi.ua.

Белошицкая Оксана Константиновна – Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Факультет биомедицинской инженерии, старший преподаватель кафедры биомедицинской инженерии; г. Киев, проспект Победы, 37 e-mail: o.biloshytska@kpi.ua.

Biloshytska Oksana – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Faculty of Biomedical Engineering, Senior Lecturer of department of Biomedical Engineering; Peremogy ave., 37, Kyiv; e-mail: o.biloshytska@kpi.ua.