

Д. О. Шевага; О. К. Городецька, к. т. н., доц.; Л. М. Добровська, к. п. н., доц.

ПРОГРАМНИЙ ДОДАТОК МОНІТОРИНГУ РІВНЯ СТРЕСУ НА ОСНОВІ МОДЕЛЕЙ КЛАСИФІКАЦІЇ

У статті наведено дослідження рівня стресу. Після пандемії COVID-19 та постійного перебування у вимушеній ізоляції, рівень стресу став вищим через зростання тривоги. Тому велику цікавість викликає дослідження основних механізмів стресу та моніторинг різних біофізіологічних і біохімічних реакцій організму на стрес. Надійний біомаркер або індикатор стресу міг би забезпечити точний моніторинг стресу, потенційно дозволяючи запобігти патологічним станам на ранніх стадіях. Тривалий стрес може мати серйозні наслідки для здоров'я. Тому здатність визначати, коли людина перебуває в стані стресу, може бути дуже корисним для запобігання проблем зі здоров'ям, особливо у пацієнтів із суїцидальними думками.

У цій роботі наведені результати дослідження моніторингу рівня стресу шляхом використання в якості прогнозування моделі класифікації, а в якості біосигналу – варіабельність серцевого ритму (HRV) від датчиків електрокардіографії. Проведено кореляцію всіх змінних для того, щоб у навчанні моделей брали участь лише ті, змінні, які мають високу кореляцію зі стресом. Для досягнення поставленої задачі використано методи: штучної нейронної мережи, k -найближчих сусідів (KNN), випадкового лісу, дерева рішень. Модель класифікації випадкових лісів (random forest) отримала найвищий показник точності прогнозування наявності чи відсутності стресу у людини – 98 %. На основі цієї моделі розроблено програмний додаток на мові програмування R з інтерфейсом користувача, який дозволяє завантажити дані електрокардіограми та отримати висновок щодо рівня стресу. За допомогою додатку користувач може здійснювати моніторинг рівня особистого стресу та вести здоровий спосіб життя.

Ключові слова: моніторинг стресу, електрокардіографія, машинне навчання, варіабельність серцевого ритму, алгоритми класифікації, біосигнали, дерево рішень, випадковий ліс.

Постановка проблеми

Дослідження в області стресу показують, що стрес може викликати порушення серцевого ритму, що може призвести до аритмії та завдати шкоди як фізичному, так і психічному здоров'ю [1]. Фактично, хронічний стрес може збільшити шанси на розвиток депресії, появу інсульту, діабету та навіть раку. Стрес впливає на всі системи організму, включаючи кістково-м'язову, дихальну, серцево-судинну, ендокринну, шлунково-кишкову, нервову та репродуктивну системи. Крім того, стрес визначали як одну з головних причин автомобільних аварій, що призводить до високих показників смертності [2]. З цих причин важливо розробити додаток, який зможе надійно виявляти та вимірювати людський стрес неінвазивно. У зв'язку з тим, що стрес у кожної людини проявляється по-різному, необхідно створити модель та програмний додаток, який буде персоналізовано для людини оцінювати її рівень стресу враховуючи її вхідні показники. В якості параметру для оцінки рівня стресу було обрано показники варіабельності серцевого ритму, частота серцевих скорочень (ЧСС), та модифікації RR інтервалів від датчиків електрокардіографії (ЕКГ), які підтверджуються багатьма дослідженнями [3 – 5]. Вибір цих метрик було відібрано при створенні бази даних.

Аналіз останніх досліджень та публікацій

Зайвий стрес може спричинити серйозні проблеми та збільшити ризик захворювання. Сьогодні існують дослідження щодо ідентифікації стресу. У роботі [6] автори представили базу даних DRIVEDB і дослідили кілька методів кластеризації, включаючи класифікацію

SVM-RBF, KNN і RBF. Найкращі результати були отримані з точністю 83 %, використавши класифікатор SVM-RBF [6]. Автори [7] визначали рівень стресу за допомогою біологічних сигналів на основі датчиків, таких як частота серцевих скорочень, параметри дихання. Методи KNN і SVM використовувалися для класифікації та були отримані високі показники точності (табл. 1). Інші дослідники аналізували стрес зайнятих людей за допомогою різних датчиків тіла, датчиків Kinect 3D, відеоспостереження, а індивідуальні рівні стресу класифікує на основі алгоритму SVM, KNN. Це дослідження проводилось на основі даних SWELL-KW, побудована модель має точність 92,75 %. У табл. 1 продемонстровані зведені дані останніх досліджень щодо класифікації стресу.

Таблиця 1

Зведена таблиця останніх досліджень щодо класифікації стресу

Authors	Dataset Used	Classifier	Classification	Accuracy
Adnan Ghaderi et.al.	PHYSIONET	KNN, SVM	Low, Medium, High	100 sec & 200 sec- 98.41% 300 sec- 90%
Nermine Munla et.al.	DRIVEDB	SVM- RBF, KNN, RBF	Normal, Stressed	Highest accuracy SVM- RBF-83%
Sriramprakash. S et.al.	SWELL-KW	SVM	Stressed, Normal	92.75%
Ravinder Ahuja et.al.	206 students of JIIT Noida	Random forest, Naïve Bayes, SVM, KNN	Highly stressed, Stressed, Normal	Highest accuracy SVM- 85.71%
Md Fahim Rizwan et.al.	PHYSIONET	SVM	Stressed, Non- Stressed	98.6%
Cheng-Ping Hsieh	WESAD	XGBoost	Stressed, Non-Stressed	92.38%
S. M. Chaware et.al.	Information extraction from Facebook attributes	TVSM	Positive, Negative	84.2%

Мета дослідження

Метою дослідження є підвищення точності прогнозування наявності стресу шляхом побудови класифікаторів на основі реалізації штучної нейронної мережі, використання методу k-найближчих сусідів, випадкового лісу та розробки програмного додатку, що дозволило оцінювати ймовірний рівень стресу, де для прогнозування моделі класифікації, використовуються результати оброблення біосигналів ЕКГ, а саме значення варіабельності серцевого ритму.

Матеріали та методи дослідження

Набір даних на якому базується це дослідження містить індекси варіабельності серцевого ритму, обчислені на основі мультимодального набору даних SWELL (SWELL-KW) для дослідження стресу. SWELL був створений дослідниками з Інституту обчислювальної техніки та інформаційних наук Університету Радбуда. Це результат експериментів, проведених на 25 суб'єктах, які виконують типову офісну роботу. Всього налічується 369288 випадків, які були досліджені на стресовий стан. Суб'єкт проходив через типові робочі стреси, такі як відволікання на надходженні листи через електронну пошту та тиск, щоб завершити роботу вчасно. Кожен учасник пройшов три різні умови роботи:

1. Без стресу (no stress): учасникам дозволяється працювати над завданнями стільки, скільки їм потрібно, максимум 45 хвилин, але вони не знають про максимальну тривалість

своїх завдань.

2. Нестача часу (time pressure): протягом цього часу час виконання завдання скоротився до 2/3 часу, який учасник витратив у нейтральному стані.

3. Відволікання (interruption): учасники отримали вісім електронних листів під час виконання призначених завдань. Деякі електронні листи мали відношення до їхніх завдань – і від учасника просили вжити певних дій – тоді як інші просто не стосувалися їх завдань.

Дані містять різні атрибути, отримані з сигналів, вимірюваних за допомогою ЕКГ, записаних для різних осіб, які мають різний пульс на момент проведення вимірювання. Ці різні особливості впливають на частоту серцевих скорочень у певний момент часу для людини.

Загалом визначено 36 різних ознак для класифікації рівнів стресу, серед яких ЧСС, інтервали RR та їх модифікації (медіана, стандартне відхилення, середньоквадратичний корінь, відсоток послідовних інтервалів тощо).

У наборі даних переважала кількість випадків, коли людина знаходилась не в спокої (рис. 1).

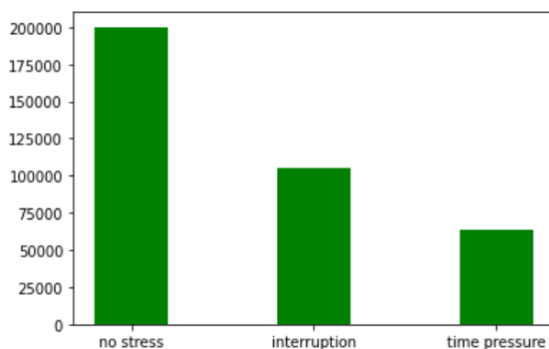


Рис. 1. Розподіл за рівнем стресу

Досліджено чотири алгоритми класифікації: метод k-найближчих сусідів, штучну нейронну мережу, випадковий ліс, дерево рішень.

Нейронні мережі, також відомі як штучні нейронні мережі (ANN) або імітовані нейронні мережі (SNN), є підмножиною машинного навчання та основою алгоритмів глибокого навчання. Штучні нейронні мережі складаються з вузлових шарів, що містять вхідний рівень, один або більше прихованих шарів і вихідний рівень. Кожен вузол, або штучний нейрон, з'єднується з іншим і має відповідну вагу та поріг [8].

Метод k-найближчих сусідів – це техніка машинного навчання, яка підпадає під контрольоване навчання. Алгоритм припускає, що подібні речі існують дуже близько одна до одної. Вибір правильного значення K є найважливішим завданням. Для заданих даних, які невідомі машині, ми можемо перевірити з різними значеннями K, щоб отримати зменшену помилку для кращого прогнозування. Спочатку вибирається кількість K сусідів, потім обчислюється евклідова відстань, і відповідно до розрахунку виконується вибір k-найближчих сусідів. Підраховується кількість точок даних у кожній категорії. Для категорії з максимальним сусідом призначаються нові точки даних, після чого модель готова [9].

Дерево рішень – це організований у вигляді дерева класифікатор, де кожен внутрішній вузол відповідає одній з вхідних змінних, гілки посилаються до вказівок щодо прийняття рішень, а кожен кінцевий вузол звертається до результату [10]. У дереві рішень є два вузли: вузол прийняття рішень і листовий вузол. Вузли рішень використовуються для ухвалення будь-якого рішення та мають різні гілки, тоді як листові вузли є результатом цих рішень і не містять подальших гілок.

Випадковий ліс – популярний ансамблевий інструмент машинного навчання на основі дерев рішень, який має високу адаптивність до даних. Цей метод зменшує ступінь

перенавчання шляхом поєднання кількох оцінювачів перенавчання (тобто дерев рішень) для формування ансамблевого алгоритму навчання [9]. Кожне дерево рішень може отримати відповідний результат класифікаційного рішення. Використовуючи результати голосування кожного дерева рішень у лісі, категорія зразка, що підлягає перевірці, визначається відповідно до принципу меншості, яка підкоряється більшості. Категорія з більш високими голосами в усіх деревах рішень визначається як остаточний результат.

Результати дослідження

Для подальшого дослідження проведено кореляцію всіх змінних для того, щоб у навчанні моделей брали участь лише ті, змінні, які мають високу кореляцію зі стресом. Оцінивши матрицю кореляції, було обрано наступні змінні для побудови моделей класифікації: середнє значення інтервалів RR, медіана інтервалів RR, співвідношення стандартного відхилення інтервалів RR до середньоквадратичного кореня послідовних різниць інтервалів RR, медіана відносних інтервалів RR, співвідношення стандартного відхилення інтервалів RR до середньоквадратичного кореня послідовних різниць інтервалів RR для відносних різниць інтервалів RR, абсолютна потужність діапазону дуже низьких частот (VLF), перетворення головного компонента VLF.

Проведено нормалізацію методом min-max, оскільки він є однією з найпоширеніших способів нормалізації даних. Для кожної функції її мінімальне значення перетворюється на 0, максимальне значення перетворюється на 1, а кожне інше значення перетворюється на десяткове значення між 0 і 1.

Для отримання об'єктивної моделі кожен вибірку поділяли на навчальну (70 %) та тестову (30 %). Метрики оцінки класифікованих моделей: точність, чутливість, влучність.

Побудована штучна нейронна мережа складається з шести шарів з прямим зв'язком, у яких дані передаються від вхідного шару до вихідного без зворотного зв'язку (рис. 2). В мережі функцією активації виступає зрізаний лінійний вузол (ReLU), проте для отримання результату класифікації на вихідному шарі використано функцію активації softmax. Функцією втрат виступає категоріальна крос-ентропія, методом оптимізації – алгоритм Adam.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 7)	56
dense_1 (Dense)	(None, 35)	280
activation (Activation)	(None, 35)	0
dense_2 (Dense)	(None, 20)	720
activation_1 (Activation)	(None, 20)	0
dense_3 (Dense)	(None, 9)	189
activation_2 (Activation)	(None, 9)	0
dense_4 (Dense)	(None, 4)	40
activation_3 (Activation)	(None, 4)	0
dense_5 (Dense)	(None, 1)	5
Total params: 1,290		
Trainable params: 1,290		

Рис. 2. Архітектура побудованої штучної нейронної мережі

Для моделі класифікації KNN за основний параметр, а саме кількість сусідів, обрано значення 10 ($k = 10$). Для моделі класифікації «Дерево рішень» параметр глибини дорівнює 13, а функцією вимірювання якості розбиття обрано критерій «entropу». Параметр глибини

для моделі випадкового лісу дорівнює 12, а кількість дерев моделі дорівнює 21. Схематичне зображення архітектури методів дерева рішень та випадкового лісу наведено на рис. 3.

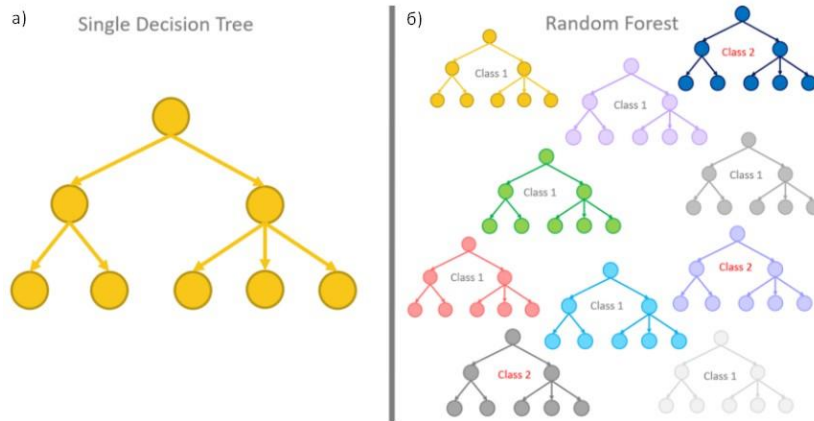


Рис. 3. Архітектура: а) дерева рішень; б) випадкового лісу

Продемонструємо результати точності для кожної побудованої моделі (рис. 4).

	Model			
	Artificial Neural Network	KNN	Random Forest	Decision Tree
Interruption	0.85	0.7	1.00	0.98
No stress	0.91	0.81	0.98	0.96
Time pressure	0.8	0.69	0.97	0.97
Accuracy	0.85	0.74	0.98	0.96

Рис. 4. Точність моделей класифікації

Як бачимо найвищий показник точності отримала прогнозна модель випадкового лісу – 0,98. Для наочності продемонструємо результати аналізу точності моделей на тестовій вибірці у вигляді стовпчикової діаграми (рис. 5).

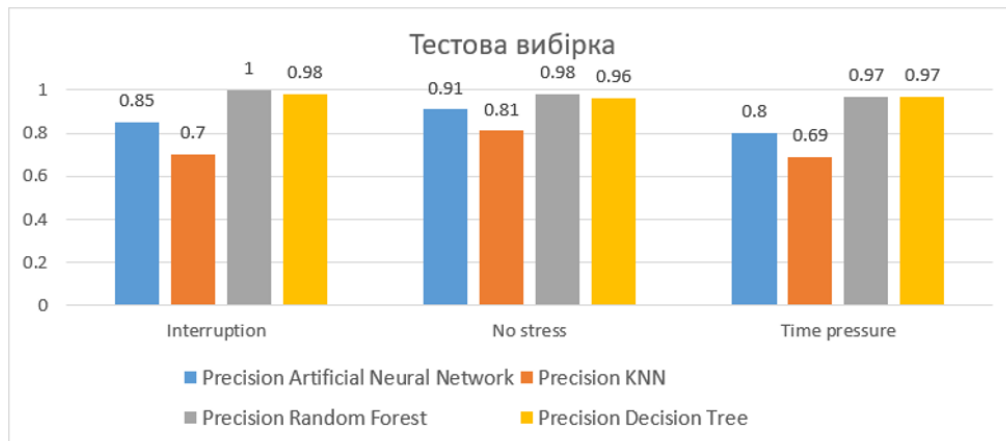


Рис. 5. Стовпчикова діаграма результатів precision

Для кращого розуміння прогнозованих моделей побудовано матрицю невідповідності – матриця NxN, яка використовується для оцінки ефективності моделі класифікації, де N –

кількість цільових класів. Матриця невідповідності для моделі класифікатора дерево рішень зображена на рис. 6.

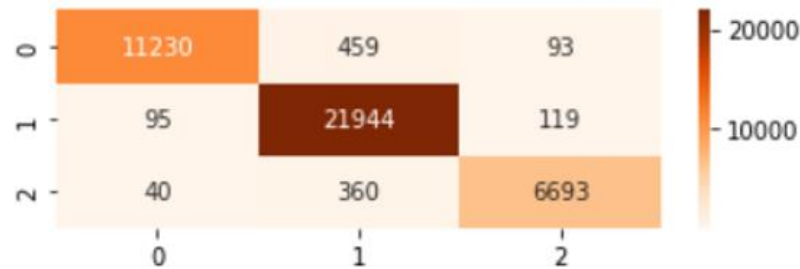


Рис. 6. Матриця невідповідності випадкового лісу на тестовій вибірці

На рис. 6 показує, що 11230 випадків вірно класифіковано в стані 0 (time pressure), 21944 вірно класифіковано в стані 1 (no stress), 6693 вірно класифіковано в стані 2 (time pressure). Загалом модель зробила 1165 хибних тверджень та вірно класифікувала 39868 випадків.

Модель класифікації випадковий ліс отримала найвищі показники точності, порівняно з іншими побудованими аналогами, тому саме вона використана для розробки програмного продукту, який допоможе користувачам керувати рівнем особистого стресу та вести здоровий спосіб життя.

За допомогою мови програмування R розроблено програмний додаток моніторингу рівня стресу. Цей додаток є кросплатформенним та може працювати як на операційній системі Windows, так і на операційній системі Linux. Дані зберігаються у вбудованій базі даних MySQL. Додаток було створено з сучасними вимогами до UI/UX [11], максимально простим і зрозумілим. Для роботи з програмою необхідно завантажити у додаток дані ЕКГ пацієнта. Вхідні дані пацієнта повинні бути записані у форматі CSV. Для моніторингу рівня стресу достатньо завантажити спостереження за одну – дві хвилини, після чого програмне забезпечення, використовуючи модель класифікації випадкового лісу, видає результат та точність зробленого передбачення.

У програмі є 4 сторінки, перша з яких – «Insert data» відповідає за завантаження даних пацієнта для виявлення наявності або відсутності стресу. Після завантаження датасету, користувач отримує результат роботи програми, а саме висновок щодо приналежності до одного з класів: no stress, interruption, time pressure та кругову діаграму його станів в період часу, який був завантажений у датасеті (рис. 7).

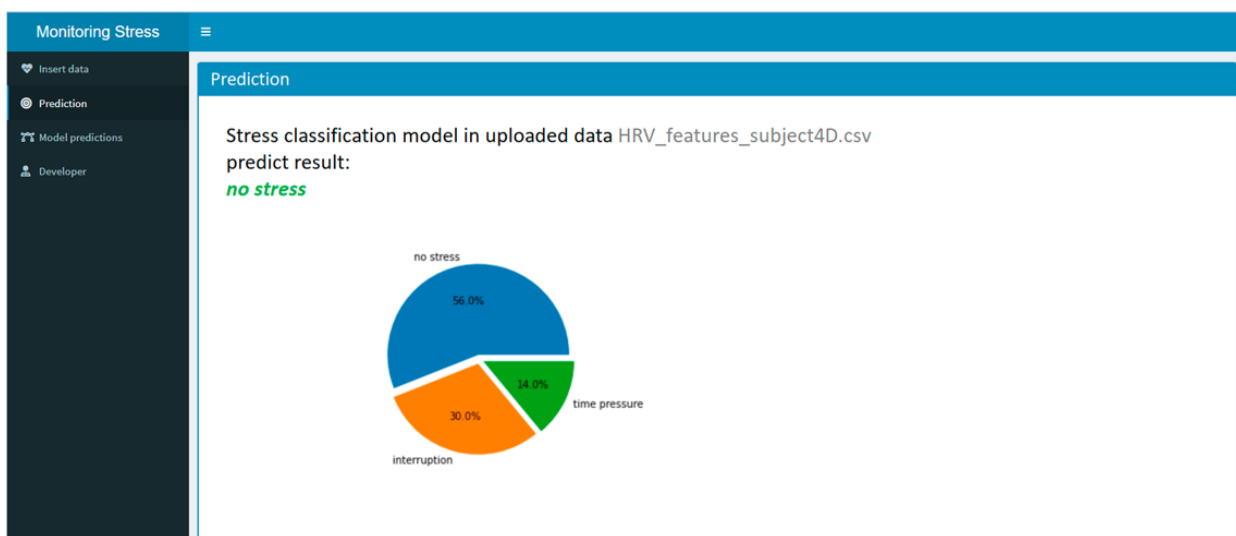


Рис. 7. Результат роботи програмного додатку

Висновки

На основі бази даних SWELL-KW проаналізовано 369288 випадків, які були досліджені на стресовий стан та побудовано 4 моделі класифікації, які дозволяють спрогнозувати стан стресу з високою точністю. Побудовані класифікатори для прогнозування наявності стресу: штучна нейронна мережа, метод k-найближчих сусідів, випадковий ліс, дерево рішень. У результаті отримали модель випадкового лісу з точністю 98 % прогнозування наявності чи відсутності стресу у людини, що є точнішим результатом порівняно з роботою Sriramprakash. S, де точність класифікованої моделі склала 92,75 %. На основі цієї моделі розроблене програмне забезпечення з інтерфейсом користувача, яке дозволяє завантажити дані електрокардіограми та отримати висновок щодо рівня стресу.

СПИСОК ЛІТЕРАТУРИ

1. Medically-oriented design for explainable AI for stress prediction from physiological measurements [Електронний ресурс] / Dalia Jaber, Hazem Hajj, Fadi Maalouf, Wassim El-Hajj // BMC Med Inform Decis Mak. – 2022. – Режим доступу до ресурсу: <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-01772-2>.
2. Choi M. Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness / M. Choi, G. Koo, M. Seo, SW. Kim // IEEE Trans Instrum Meas. – 2017. – № 67. – С. 34 – 45.
3. Thermal Comfort and Stress Recognition in Office Environment [Електронний ресурс] / K. Nkurikiyezeu, K. Shoji, A. Yokokubo, L. Guillaume // International Conference on Health Informatics. – 2019. – Режим доступу до ресурсу: https://www.researchgate.net/publication/330754493_Thermal_Comfort_and_Stress_Recognition_in_Office_Environment.
4. The SWELL Knowledge Work Dataset for Stress and User Modeling Research / Saskia Koldijk, Maya Sappelli, Suzan Verberne et al.]. // ICMI '14: Proceedings of the 16th International Conference on Multimodal Interaction. – 2014. – С. 291 – 298.
5. Koldijk S. Detecting Work Stress in Offices by Combining Unobtrusive Sensor / S. Koldijk, M. Neerinx, W. Kraaij // IEEE Transactions on Affective Computing. – 2018. – № 2. – С. 227 – 239.
6. Driver stress level detection using HRV analysis [Електронний ресурс] / N. Munla, M. Khalil, A. Shahin, A. Mourad // International Conference on Advances in Biomedical Engineering (ICABME). – 2015. – Режим доступу до ресурсу: <https://ieeexplore.ieee.org/document/7323251>.
7. Machine learning-based signal processing using physiological signals for stress detection [Електронний ресурс] / A. Ghaderi, J. Frounchi, A. Farnam // Iranian Conference of Biomedical Engineering (ICBME). – 2015. – Режим доступу до ресурсу: <https://ieeexplore.ieee.org/abstract/document/7404123>.
8. EEG signals classification using a new radial basis function neural network and jellyfish meta-heuristic algorithm [Електронний ресурс] / H. Rastegar, D. Giveki, M. Choubin // Evolutionary Intelligence. – 2022. – Режим доступу до ресурсу: <https://link.springer.com/article/10.1007/s12065-022-00802-2>.
9. Naser A. Classification of pleasant and unpleasant odor imagery EEG signals [Електронний ресурс] / A. Naser, O. Aydemir // Neural Computing and Applications. – 2022. – Режим доступу до ресурсу: <https://link.springer.com/article/10.1007/s00521-022-08171-8>.
10. Ekubo E. Using machine learning to predict low academic performance at a Nigerian university / E. Ekubo, B. Esiefarienrhe. // The African Journal of Information and Communication (AJIC). – 2022. – № 30. – С. 1 – 33.
11. Що таке UI/UX дизайн - з чого розпочати навчання [Електронний ресурс] // IT STEP ACADEMY. – 2022. – Режим доступу до ресурсу : <https://kiev.itstep.org/blog/what-is-ui-ux-design-where-to-start-learning>.

Стаття надійшла до редакції 14.12.2022.

Стаття пройшла рецензування 24.12.2022.

Шевага Діана Олександрівна – студент групи БС-11мп, e-mail: dinashevaha@gmail.com.

Городецька Олена Костянтинівна – кандидат технічних наук, доцент кафедри біомедичної кібернетики.

Добровська Людмила Миколаївна – кандидат педагогічних наук, доцент кафедри біомедичної кібернетики.

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».