

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ  
НТУУ «КПІ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
ДЕРЖАВНА ЕКОЛОГІЧНА АКАДЕМІЯ  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ОБОРОНИ УКРАЇНИ  
ІМЕНІ ІВАНА ЧЕРНЯХОВСЬКОГО  
ТОВ «ІТЦ ХАЙ-ТЕК БЮРО»  
КОМПАНІЯ «E-TRADE HUB LTD.»  
МІЖНАРОДНИЙ ІНСТИТУТ ЕКОЛОГІЧНОЇ БЕЗПЕКИ  
ОРГАНІЗАЦІЇ БЕЗПЕКИ СПІЛЬНОТИ ЄВРОПИ  
ЦЕНТР ЕКОЛОГО-РЕСУРСНОГО ВІДНОВЛЕННЯ ДОНБАСУ  
ГО «АСОЦІАЦІЯ ФАХІВЦІВ ЦИВІЛЬНОГО ЗАХИСТУ»

Третя міжнародна  
науково-практична конференція

**«Сучасні тенденції розвитку  
інформаційних систем  
і телекомунікаційних технологій»**

25–26 січня 2021 р.

Київ НУХТ 2021

Наукові праці Третьої міжнар. наук.-практ. конф. «Сучасні тенденції розвитку інформаційних систем і телекомунікаційних технологій», 25–26 січня 2021 р. (Київ, Україна). – К. : НУХТ, 2021. – 181 с.

У працях конференції наведено доповіді за напрямками:

- світові тенденції в розробленні інформаційних систем і телекомунікаційних технологій;
- міжнародні стандарти в галузі інформаційних і телекомунікаційних технологій та кіберзахисту;
- розвиток освіти і науки в галузі інформаційних і телекомунікаційних технологій та кіберзахисту;
- інтернет речей та розвиток його технологій для безпечного суспільства;
- моделювання та симуляція стихійних лих, надзвичайних ситуацій і реагування на них;
- досвід використання інформаційних технологій, безпілотних літальних апаратів і роботів для моніторингу навколишнього середовища, попередження й ліквідації надзвичайних ситуацій природного і техногенного походження;
- неурядові та громадські організації у сфері цивільного захисту.

Праці конференції будуть корисні науковим та інженерно-технічним працівникам, студентам ЗВО та всім, хто цікавиться сучасними інформаційними системами та телекомунікаційними технологіями.

**Подано в авторській редакції.**

**ISBN 978-83-956296-3-1**

**© НУХТ, 2021**

<i>13. Власенко Л. О., Мурга І. В.</i> Моделювання бази даних автоматизованої системи формування документів для вступних випробувань на основі семантичних підходів.....	55
<i>14. Ворона Ю. С., Сірик А. О., Чумаченко С. М.</i> Системний підхід до розроблення загальної моделі ризику виробничого травматизму на підприємстві харчової промисловості.....	58
<i>15. Галайда Ю. Ю.</i> Методологічні засади портфельно-орієнтованого управління.....	59
<i>16. Гладка Ю. А., Назаренко Є. О.</i> Аналіз застосування технологій штучного інтелекту в кібербезпеці.....	64
<i>17. Гладкий Я. В., Гладка М. В., Лісневський Р. В., Костіков М. П.</i> Методи кластеризації як інструмент налаштування нечіткої моделі.....	66
<i>18. Євсенко Д. О.</i> Програмно-апаратний комплекс для моніторингу радіаційної обстановки в зоні впливу радіаційно-небезпечних об'єктів.....	71
<i>19. Жеваго О. О., Шинкаренко В. І.</i> Аналіз та вдосконалення навчального процесу з програмної інженерії з використанням методів Process Mining.....	73
<i>20. Загоровська Л. Г., Петрик В. В.</i> Кластерний аналіз даних у маркетинговій діяльності торговельного підприємства.....	75
<i>21. Загоровська Л. Г., Химич М. М.</i> Інформаційне забезпечення задач аналізу та прогнозування екологічних показників м'ясокомбінату.....	79
<i>22. Іванов І. І., Тімінський О. Г.</i> Моделі управління проектом створення ціннісно-орієнтованої відеогри.....	82
<i>23. Іванченко Н. О., Подскребко О. С.</i> Застосування інтернету речей у сільському господарстві.....	86
<i>24. Карпенко М. І., Чумаченко С. М., Мошенський А. О.</i> Розроблення системи комплексного біоімпедансного діагностування.....	90
<i>25. Кашапова Л. Р., Хлевний А. О.</i> Застосування проектного підходу до створення бізнесу з продажу тортів на замовлення.....	92

УДК 004.81

**МЕТОДИ КЛАСТЕРИЗАЦІЇ ЯК ІНСТРУМЕНТ  
НАЛАШТУВАННЯ НЕЧІТКОЇ МОДЕЛІ**

**Гладкий Я. В.**

*Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна*

**Гладка М. В., Лісневський Р. В., Костіков М. П.**

*Київський національний університет імені Тараса Шевченка, Київ, Україна  
E-mail: MikolaszK@gmail.com*

**Clusterization Methods as a Tool of Fuzzy Model Adjustment**

*The paper considers creating a radial basis function (RBF) network with evolutionary optimization. The subject of study is evolutionary optimization of neurons' hyperparameters within a network of radially basis functions with the help of a genetic algorithm. An RBF network was built in order to classify the data. The object of study is a dataset of hepatocellular carcinoma*

(HCC) that was collected at a university hospital in Portugal. It contains the real clinical data of 165 patients with HCC diagnosis. This dataset contains a classifying attribute: whether the patient has survived within 1 year. This attribute is used directly for classifying the data by the RBF network.

RBF-мережі — вид великого класу під назвою нейронні мережі, які є сучасним інструментом розв’язання задач будь-якої складності. Усі задачі, які можуть розв’язувати нейронні мережі, так чи інакше пов’язані з навчанням. Серед основних областей їх застосування: прогнозування, прийняття рішень, розпізнавання образів, оптимізація, аналіз даних. Мережі RBF являють собою локальні апроксиматори для нелінійного відображення вводу-виводу. Їх основними перевагами є короткий етап навчання та знижена чутливість до порядку подання навчальних даних. Однак у багатьох випадках ми виявляємо, що плавне відображення досягається лише тоді, коли кількість радіальних базисних функцій, необхідних для охоплення вхідного простору, стає дуже великою. Це заважає здійсненню багатьох практичних застосувань [2].

Мережа RBF має лише один прихований шар, а кількість базових функцій та їх форма орієнтована на проблеми та може бути визначена в мережі під час навчального процесу. Кількість нейронів у вхідному шарі дорівнює розмірності вектора ознак. Так само кількість вузлів у вихідному рівні відповідає кількості класів [3].

Однією з головних перешкод на шляху широкого використання штучних нейронних мереж є складність адекватного визначення значень для їх вільних параметрів, так званих гіперпараметрів. Модель Еріка Майларда та Дідьє Гуеріота, французьких дослідників, які у 1997 році запропонували свій спосіб застосування генетичного алгоритму для оптимізації гіперпараметрів мережі [4] дозволяє розробити алгоритм перевірки наслідковості за ораними параметрами.

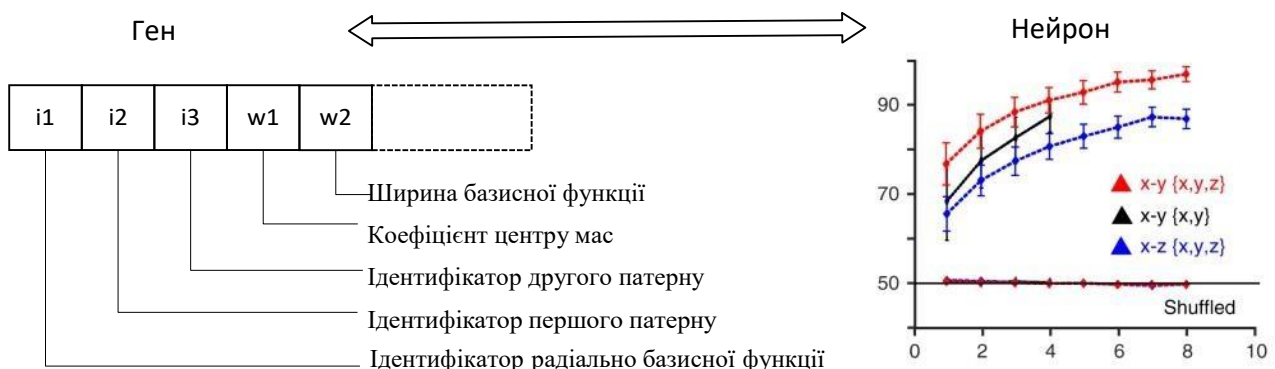


Рис. 1. Відповідність гену та параметрів нейрону, який він визначає

Для еволюційної оптимізації було застосовано модель Майларда і Геріота, суть якої заключається в наступному: еволюція RBF-мережі відбувається шляхом генетичної модифікації параметрів для кожного окремого нейрону. Один геном має рівно стільки генів, скільки нейронів у прихованому шарі. Кожен ген містить інформацію про 5 параметрів (рис. 1):

- ідентифікатор радіально базисної функції;

- ідентифікатор першого патерну;
- ідентифікатор другого патерну;
- коефіцієнт центру мас;
- ширина базисної функції.

Терміном патерн у галузі машинного навчання називається пара вхід-вихід з тренувального набору даних [5]. Набір даних ГЦК був отриманий в університетській лікарні в Португалії і містить кілька демографічних факторів, факторів ризику, лабораторних та загальних особливостей виживання 165 реальних пацієнтів з діагнозом ГЦК. Набір даних містить 49 факторів, вибраних відповідно до рекомендацій клінічної практики EASL-EORTC (Європейська асоціація з вивчення печінки — Європейська організація з досліджень та лікування раку), які є сучасними методами управління НСС.

Це різномірний набір даних, який містить 23 кількісні змінні та 26 якісних змінних. Загалом відсутні дані становлять 10,22% від усього набору даних, і лише вісім пацієнтів мають повну інформацію з усіх полів (4,85%). Цільовими змінними є виживання за 1 рік і кодувалася як двійкова змінна: 0 (вмирає) та 1 (живе). Певний ступінь дисбалансу між класами також присутній (63 випадки, позначені як «мертвий», а 102 як «живий»).

Для навчання РБФ-мережі було обрано власноруч 10 факторів, які потенційно можуть мати найбільший зміст:

- 1) відсоток феритину у крові;
- 2) відсоток насичення киснем кров'ю;
- 3) відсоток заліза у крові;
- 4) прямий білірубін;
- 5) основний розмір утворення;
- 6) кількість складових утворення;
- 7) відсоток креатиніну;
- 8) кількість загальних білків;
- 9) кількість лужної фосфатази;
- 10) кількість гамма-глутамілтрансферази.

Дані розбиваються на атрибути ознак та класифікаційний атрибут, що служитиме значенням виходу мережі. Пропуски даних заміщуються середнім по кожному атрибуту, нормалізуються, розбиваються на тренувальний та тестовий набір даних у співвідношенні 2:1. Створення РБФ-функцій, які використовуватимуть нейрони, забезпечує можливість створення моделі для тренування [7]. Клас РБФ-мережі класифікуватиме набір даних між двома класами: «Вжив» або «Помер». Цей клас містить метод для декодування параметрів нейрону з кожного гену геному. Усі декодовані нейрони стають складовими мережі. Також реалізований метод, що знаходить синаптичні ваги між прихованим та вихідним шаром методом найменших квадратів [6] та вираховує прогнозований вихід на вхідному наборі даних.

Створення цього класу супроводжується усіма операторами генетичного алгоритму та головним методом старту еволюції. Саме цей метод генерує покоління з чисельністю популяцій у розмірі 10 штук. Кожна популяція містить

геном, що складається з 10 генів, який характеризує кожен нейрон РБФ-мережі. Далі вираховується тренувальна функція для кожного геному з популяції, що кількісно виражає, наскільки погано проявляє себе мережа з такою параметризацією нейронів, тобто що більше значення функції, то гірший конкретний набір генів. Два найкращі геноми залишаються на наступне покоління та з допомогою операторів кросоверу та мутації породжуються інші геноми покоління. Усього було породжено 100 поколінь. Кількість нейронів, що містить РБФ мережа, та які розміщені у прихованому шарі було визначено як 10. Вихідний перцептрон мережі один, і він набуває цільового значення 0 або 1. Архітектура РБФ-мережі зображена на рис. 2.

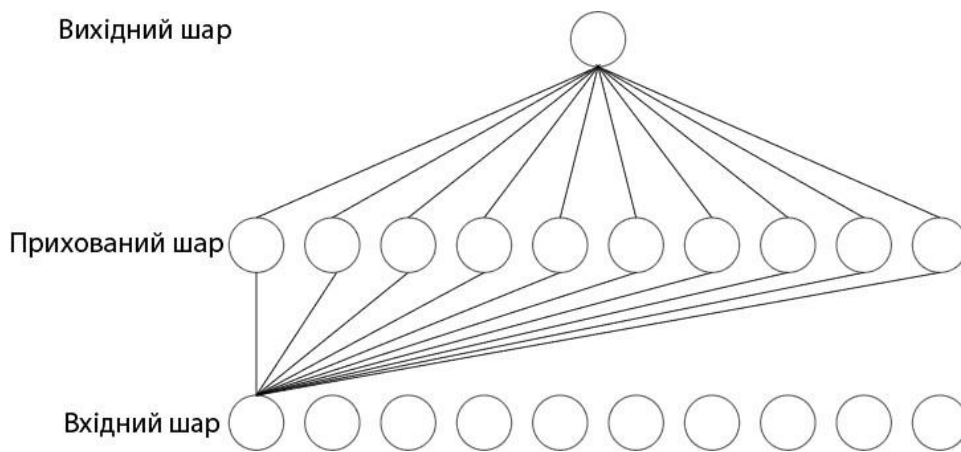


Рис. 2. Архітектура РБФ-мережі

На кожному поколінні вираховувалась тренувальна функція, що в загальному в процесі еволюції повинна була б зменшуватись. На графіку, зображеному на рис. 3, прослідковується ця тенденція.

Також, в процесі роботи алгоритму записувалась значення фітнес-функції найкращого геному у поколінні. Еволюційні зміни найкращого геному продемонстровані на рис. 4.

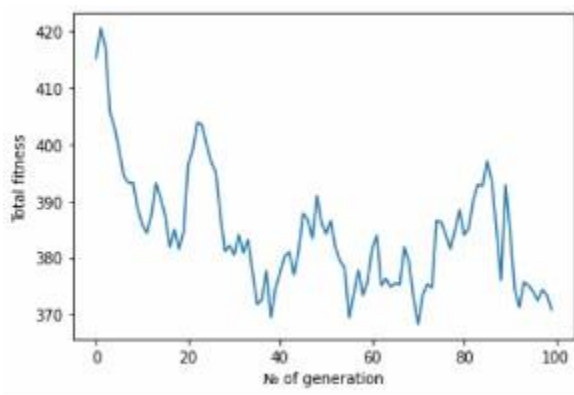


Рис. 3. Залежність значення загальної функції популяції від номеру покоління

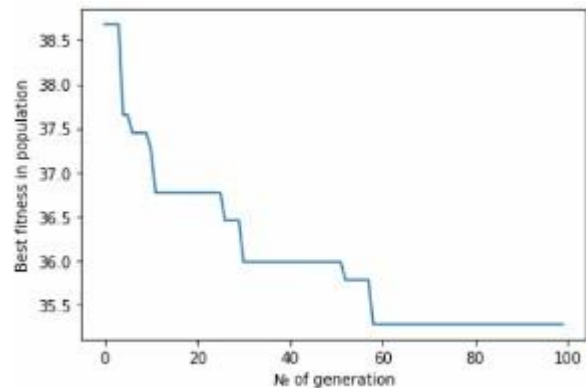
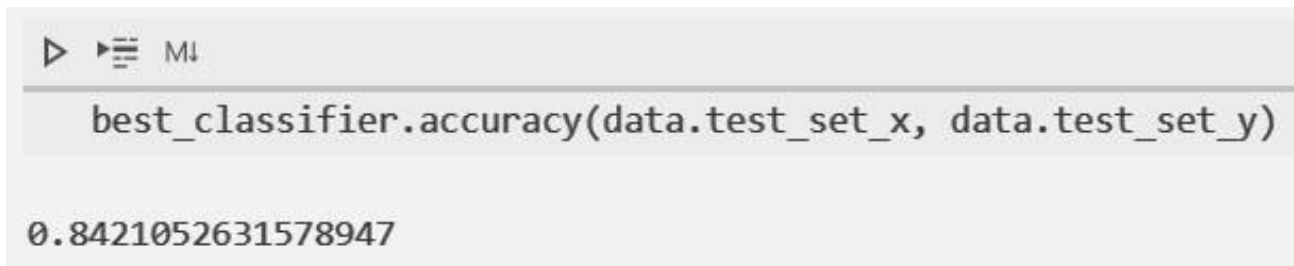


Рис. 4. Залежність значення функції з найкращим геномом у популяції від номеру покоління

По завершенні еволюції фінальний класифікатор із найкращими параметрами для РБФ-мережі був протестований на тестовому наборі даних. Знімок екрану із значенням точності класифікатора зображено на рис. 5. Як видно, 84,2% тестових даних еволюціонована модель класифікує правильно.



```
▶ ▶≡ MI
best_classifier.accuracy(data.test_set_x, data.test_set_y)
0.8421052631578947
```

Рис. 5. Точність класифікатора після еволюції у 100 поколінь

У результаті виконання дослідження була отримана еволюціонована модель, що розвинулась до точності у понад 84% за 100 поколінь. Вона вирішувала задачу класифікації між двома класами «живий» і «мертий» на наборі даних реальних 165 пацієнтів з діагнозом гепатоцелюлярної карциноми.

### Література

1. Santos M. S. *et al.* (2021) *Hepatocellular Carcinoma Dataset (HCC dataset)*, URL: <https://www.kaggle.com/mrsantos/hcc-dataset?select=hcc-data.arff>.
2. Пегат А. (2009) *Нечёткое моделирование и управление*, Москва: БИНОМ. Лаборатория знаний, 798 с.
3. Schmid A. (2014) *Pattern Recognition and Signal Analysis in Medical Imaging*, Academic Press, part 5, 372 p.
4. Maillard E., Guériot D. (1997) 'RBF Neural Network, Basis Functions and Genetic Algorithm', *International Conference on Neural Networks*, vol. 4.
5. *Machine Learning Glossary and Terms* [e-book] (2021) URL: <https://deeptai.org/machine-learning-glossary-and-terms>.
6. Molugaram K., Rao G. (2017) *Statistical Techniques for Transportation Engineering*, Butterworth-Heinemann, 554 p.
7. Леоненков А. (2005) *Нечёткое моделирование в среде MATLAB и fuzzyTECH*, СПб: БХВ-Петербург», 725 с.