

DOI: [https://doi.org/10.25140/2411-5363-2025-3\(41\)-185-192](https://doi.org/10.25140/2411-5363-2025-3(41)-185-192)

УДК 004.85:004.77

**Денис Сергійович Гончаров**

аспірант кафедри комп'ютерної інженерії

Чорноморський національний університет імені Петра Могили (Миколаїв, Україна)

E-mail: [honcharov.denys@chmnu.edu.ua](mailto:honcharov.denys@chmnu.edu.ua). ORCID: <https://orcid.org/0009-0004-1200-6677>**ДІАГНОСТИКА РАКУ ГРУДЕЙ: ІОТ-СИСТЕМА З FPGA  
ТА WEKA-АНАЛІТИКОЮ**

У сучасному світі все більше приділяється увага науковцям аналізу раку молочної залози. У статті проводиться огляд алгоритмів стиснення зображень, отриманих з ПЛІС систем для подальшої роботи у застосунку WEKA. Запропоновано можливості для ранньої діагностики раку молочної залози. Системи, які поєднують інтернет речі (ІоТ), програмовані логічні схеми (ПЛІС) та програму для аналізу даних, дозволять проводити щоденний моніторинг стану здоров'я та діагностувати можливі захворювання на ранній стадії. Також описано, як стиснення медичних зображень впливає на точність їхнього аналізу.

**Ключові слова:** ПЛІС; WEKA; інформаційні технології; стиснення зображень; jpeg-ls.

Рис.: 5. Табл.: 3. Бібл.: 8.

**Актуальність теми дослідження.** Рак молочної залози залишається найпоширенішим онкологічним захворюванням у жінок. Його рання діагностика дає можливість успішно діагностувати пухлини на ранній стадії їхнього утворення. Для підвищення ефективності цього процесу сучасна медицина потребує поєднання такої загальноприйнятої діагностики, як мамографія, з носимими пристроями для щоденного моніторингу стану здоров'я жінки. Тому сьогодні дуже важливо поєднувати нові технології, такі як спеціальні прискорювачі й розумні програми для аналізу даних, для створення безпечніших медичних рішень.

**Постановка проблеми.** Звичайні методи діагностики мають кілька недоліків. Наприклад, такий метод діагностики, як мамографія, зазвичай роблять лише раз на рік. Таким чином, залишається великий проміжок часу, під час якого рак може розвиватися непомітно. З іншого боку, сучасні системи, що використовують штучний інтелект, потребують великих обчислювальних потужностей. Якщо оптимізувати таку систему таким чином, щоб вона потребувала менших потужностей, це дозволить використання простих пристроїв для моніторингу за станом здоров'я у реальному часі.

**Аналіз останніх досліджень.** Сучасні дослідження присвячені таким напрямкам, як інтернет речей (ІоТ), ПЛІС як прискорювач та використання WEKA (Waikato Environment for Knowledge Analysis) для аналізу даних.

У статті [1] автори досліджують виявлення раку молочної залози за допомогою термографічних зображень. Таким чином, вони описують популярний метод візуалізації діагностики раку молочної залози – мамографію, використання якого не дозволяє виявляти захворювання на ранній стадії. Автори пропонують використання інфрачервоної термографії як нову технологію для виявлення злоякісних захворювань на ранній стадії.

Автори статті [2] пропонують використовувати для постійного моніторингу здоров'я рішення на основі FPGA, а саме моделі Xilinx Spartan-7, яка за результатами їх дослідження продемонструвала високу енергоефективність, що дозволило подовжити час роботи системи та термін служби батареї, а також провели порівняння їх системи з аналогічними, але на базі мікроконтролерів. Таким чином, автори дійшли висновку про більшу ефективність та пропускну здатність системи, що досліджувалась.

Стаття [3] присвячена методам інтелектуального аналізу, які використовуються для аналізу даних, отриманих з носимих пристроїв. У ній описується використання академічного набору даних про рак молочної залози для проведення інтелектуального аналізу даних в застосунку WEKA. Крім того, у ній описується оптимізація даних за допомогою попередньої обробки та вилучення малозначущих ознак перед початком аналізу. Результати дослідження демонструють, що класифікатор лінивого IBK k-NN може досягти точності 98 %.

У статті [4] розглядаються труднощі аналізу медичних зображень, такі як: високі обчислювальні вимоги та затримка моделей глибокого навчання у медичній візуалізації. У своїй роботі автори представляють апаратний прискорювач FPGA з використанням системи на кристалі Zynq XC7Z020.

Стаття [5] описує апаратну реалізацію двох прискорювачів для двох альтернативних алгоритмів візуалізації, які автори отримали зі специфікації SystemC. Перший пов'язаний з комунікацією, і вимагає перекриття та конвеєрної обробки, а другий – з обчисленнями та використовує складні математичні функції.

У дослідженні [6] розроблено надійний метод класифікації раку молочної залози під назвою Shepherd на основі штучного інтелекту Crow. Для вибору найкращих маршрутів автори використовували безпечну обробку маршрутів з використанням FACS.

У статті [7] автори також використовували апаратний прискорювач FPGA на кристалі Zynq XC7Z020, як і автори статті [4], але використовували з процесором ARM Cortex-A9.

У статті [8] підсумовують результати розгортання моделі на різних обчислювальних пристроях із використанням глибокого навчання з FPGA. Автори включили у свою роботу оцінку продуктивності моделі, пропускну здатність та порівняння затримки з різними розмірами пакетів. У результаті дослідники дійшли висновку, що FPGA є найбільш придатним для використання як прискорювач глибокого навчання.

**Виділення недосліджених частин загальної проблеми.** Недослідженою частиною є створення системи, яка буде поєднувати різні технології. Таке поєднання дозволить зробити діагностику доступнішою, частішою, такою, яка б не потребувала великих обчислювальних потужностей.

**Мета дослідження** – створити ефективну, портативну систему для діагностики раку молочної залози, використовуючи поєднання ПЛІС, IoT та аналізу даних за допомогою програми WEKA.

**Виклад основного матеріалу.** Програмована логічна інтегральна схема – це пристрій, який можна налаштувати навіть після його виготовлення. Це робить його дуже гнучким і дозволяє швидко змінювати або оновлювати його функції.

У запропонованій системі передбачається робота зі значними об'ємами даних, тому використання ПЛІС є ідеальною базою для її створення, оскільки дозволяє паралельно виконувати декілька операцій та швидко обробляти великі обсяги даних. Крім того, значною перевагою є гнучкість параметрів, що дає можливість налаштувати ПЛІС для виконання конкретних задач.

Таким чином, запропонована система складається з кількох основних частин: блоку збору даних від сенсорів, блоку попередньої обробки сигналів, модуля стиснення зображень, контролера інтерфейсів зв'язку та модуля передавання даних у застосунок WEKA. Для її реалізації використовується плата Xilinx Spartan-7 у середовищі Vivado Design Suite. Основна логіка описана мовою VHDL, а для взаємодії з сенсорами та передавання даних застосовано готові IP-ядра для UART та SPI. Сенсори з'єднані з ПЛІС через ці інтерфейси, а передача результатів здійснюється через модуль ESP8266. Модуль стиснення результатів реалізований за допомогою алгоритму JPEG-LS, який дозволяє зменшити обсяг медичних зображень без втрати важливих деталей.

Іншими словами, алгоритм роботи запропонованої системи працює таким чином. Коли пацієнт використовує систему, датчики збирають інформацію. Ця інформація спочатку надходить до ПЛІС, яка її обробляє. ПЛІС здатна виконувати багато операцій одночасно, що робить її ідеальною для швидкої обробки великих даних. Після цього дані відправляються до бази даних, а потім потрапляють у програму WEKA для подальшого аналізу (рис. 1).

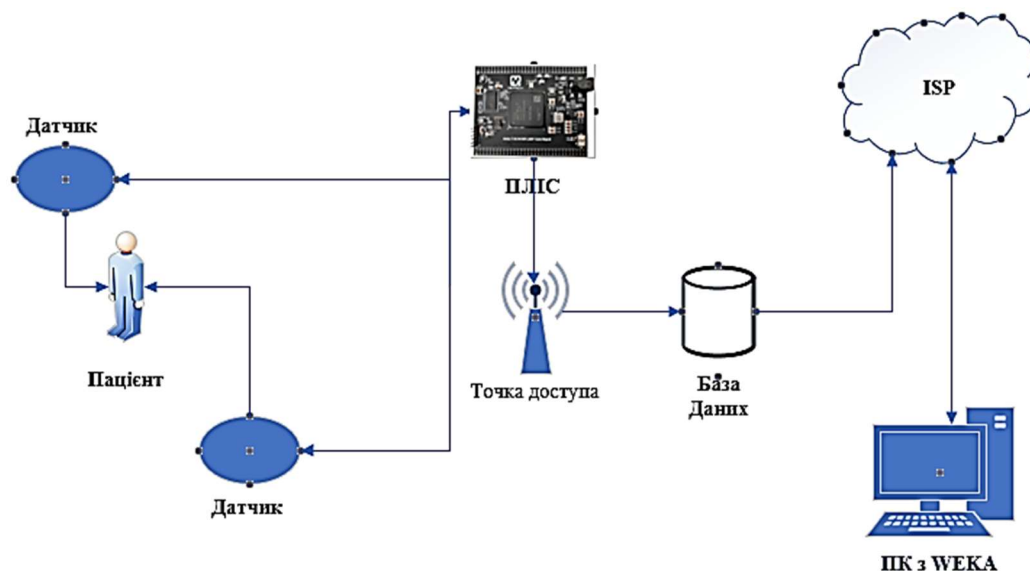


Рис. 1. Схема роботи ПЛІС системи

Джерело: розроблено автором.

Така структура робить систему швидкою, дає змогу обробляти дані паралельно й не перевантажує центральний процесор. Саме тому для цієї задачі доцільно застосовувати FPGA, а не звичайний мікроконтролер чи DPS-процесор.

#### Стиснення зображень раку молочної залози.

Існує безліч алгоритмів для стиснення зображень, таких як: jpeg, jpeg 2000, jpeg-ls, webp, deep learning lossy compression. Проте для медичних цілей краще застосовувати алгоритми, які стискають дані без втрат. Медичні зображення можуть містити дрібні структури, втрата яких може призвести до некоректного аналізу.

Таблиця 1 – Відомості про алгоритми стиснення

| Алгоритм                        | Тип стиснення                        | Переваги                                                          | Недоліки                                                                  | Використання в медицині                                              |
|---------------------------------|--------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------|
| JPEG                            | З втратами                           | Простий, швидкий                                                  | Сильні артефакти під час високого стиснення, погана робота з деталями     | Використовується нечасто через втрату дрібних деталей                |
| JPEG 2000                       | Без втрат / з регульованими втратами | Висока якість, підтримка ROI, високий динамічний діапазон         | Вища обчислювальна складність                                             | Широко використовується DICOM, особливо для медичних зображень       |
| JPEG-LS                         | Без втрат / майже без втрат          | Висока швидкість, низька складність, добре стискає гладкі області | Менш ефективний для дуже текстурованих зображень                          | Використовується у медичній візуалізації                             |
| Deep Learning Lossy Compression | З адаптивними втратами               | Адаптивне збереження важливих структур                            | Потребує великих наборів даних, складність реалізації, ще не є стандартом | Перспективний напрямок для збереження ключових деталей у зображеннях |

Джерело: розроблено автором.

Для порівняння ефективності алгоритмів було взято зображення раку молочної залози, результати якої наведені в таблиці 2.

Таблиця 2 – Стиснення зображення різними алгоритмами

| Метод               | Оригінал                                                                                               | Довжина рядка (біти) | Стиснення рядка                                                         | Довжина стисненого рядка (біти) | Коефіцієнт стиснення |
|---------------------|--------------------------------------------------------------------------------------------------------|----------------------|-------------------------------------------------------------------------|---------------------------------|----------------------|
| JPEG                | 11001100<br>10101110<br>11001010<br>11001100<br>11001010<br>11001010<br>11001010<br>1100101<br>1100101 | 71                   | 11001110 10101100<br>10101001 00110100<br>10100110 01101001<br>00101110 | 56                              | 1,26                 |
| JPEG 2000           |                                                                                                        | 71                   | 11100110 01011001<br>10100110 01100110<br>0100110110                    | 42                              | 1,69                 |
| JPEG-LS             |                                                                                                        | 71                   | 10101100<br>10100110<br>01001101<br>00101001<br>1010                    | 36                              | 1,97                 |
| Deep Learning Lossy |                                                                                                        | 71                   | 10011010<br>00110010<br>11001100<br>11000100<br>1010                    | 36                              | 1,97                 |

Джерело: розроблено автором.

### Аналіз даних за допомогою WEKA.

Для експерименту в застосунку WEKA з відкритого доступу було взято набір даних, які містять різні фізіологічні показники та їхній зв'язок з наявністю або відсутністю раку молочної залози.

У першому експерименті для побудови моделі було застосовано RandomForest, який створює рішення на основі багатьох дерев, що забезпечує високу стійкість і точність. Аналіз проводився на наборі даних із 596 прикладів і 32 атрибутами з використанням десятикратної крос-валідації для підвищення надійності оцінки.

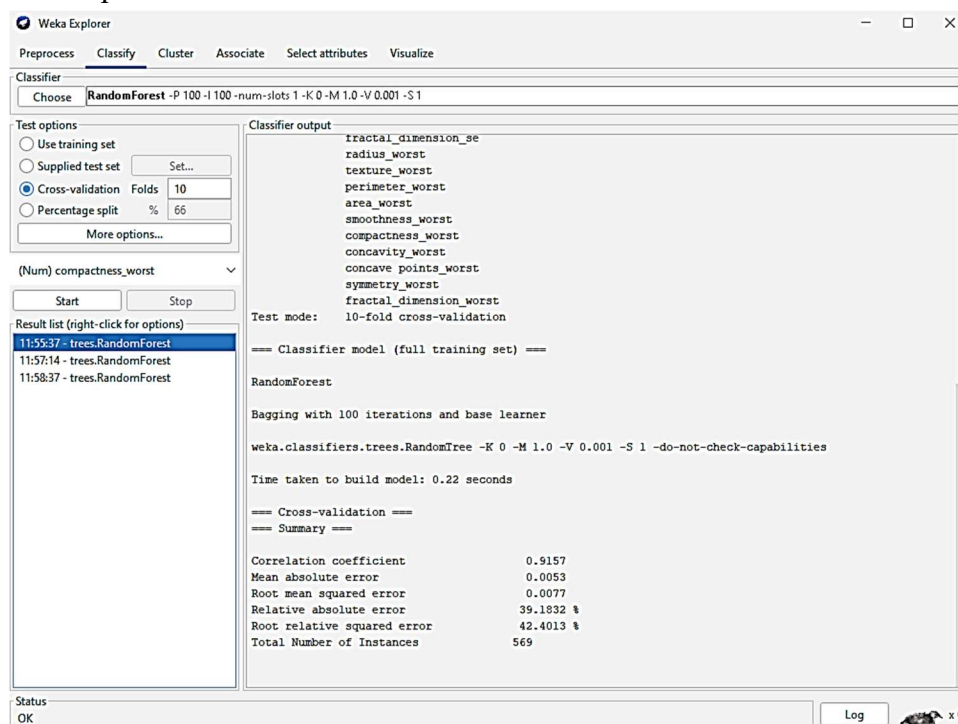


Рис. 2. Побудова моделі з 32 атрибутів

Джерело: розроблено автором.

Отримані результати свідчать про високу якість моделі. Коефіцієнт кореляції становить 0,9157, що вказує на сильний зв'язок між прогнозованими та фактичними значеннями. Середня абсолютна похибка дорівнює 0,0053, а корінь середньоквадратичної похибки – 0,0077, що демонструє низький рівень відхилення прогнозів від реальних даних. Відносні значення похибок 39,18 % і 42,40 % також підтверджують прийнятну точність при роботі з багаторівневими медичними показниками (рис. 2).

У другому експерименті був використаний набір, що містив 569 прикладів та 31 атрибут (усі, крім ідентифікатора пацієнтів). Оцінювання також проводилося за допомогою десятикратної крос-валідації. Цього разу результати аналізу показали високу точність прогнозування. Коефіцієнт кореляції склав 0,9159, що вказує на дуже сильний зв'язок між фактичними та передбаченими значеннями. Середня абсолютна похибка дорівнює 0,0052, а корінь середньоквадратичної похибки 0,0076, що свідчить про мінімальне відхилення прогнозів від реальних даних. Відносні показники похибки також залишилися низькими 38,46 % і 42,20 %, що підтверджує стабільність роботи алгоритму (рис. 3).

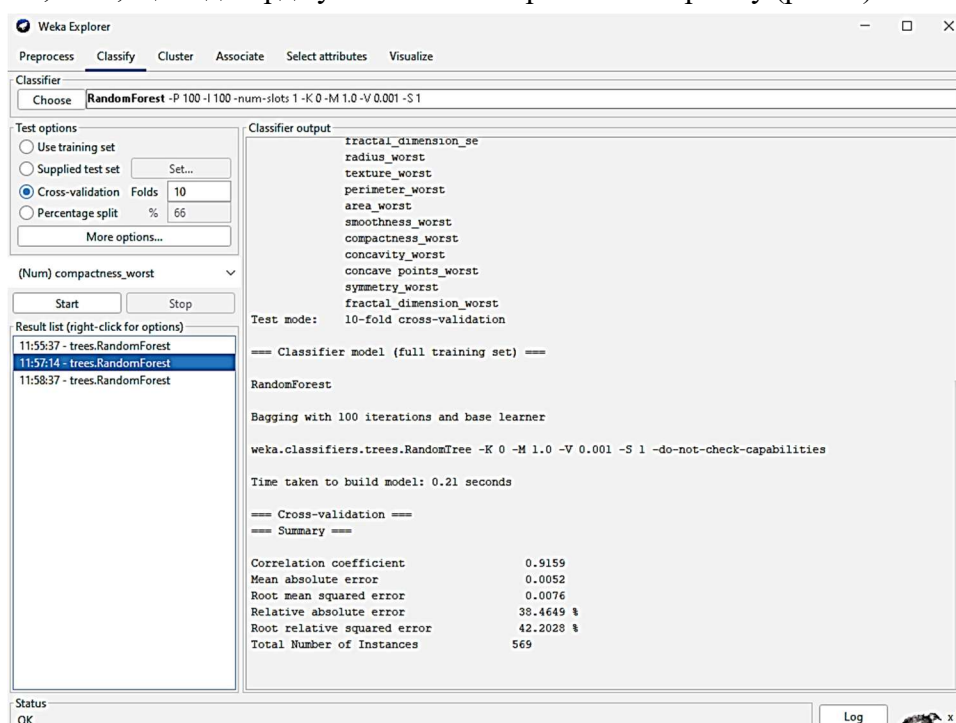


Рис. 3. Побудова моделі з 31 атрибута

Джерело: розроблено автором.

Під час виконання третього експерименту на початку роботи проведемо аналіз даних за допомогою методу CFsSubsetEval у поєднанні з пошуком Best First. Таким чином, було відібрано найбільш значущі атрибути з початкового набору, що складався з 32 показників і 596 прикладів. У результаті перевірки 295 можливих комбінацій ознак було визначено оптимальну підмножину з 6 атрибутів, яка надала високий рівень узгодженості з цільовою змінною показників придатності на рівні 0,846.

До відібраних характеристик увійшли: smoothness\_mean, fractal\_dimension\_mean, compactness\_se, fractal\_dimension\_se, smoothness\_worst та compactness\_worst (рис. 4).

Це свідчить про те, що саме параметри, пов'язані з гладкістю поверхні клітин, компактністю та фрактальними вимірами, мають найбільший вплив на визначення стану зразків.

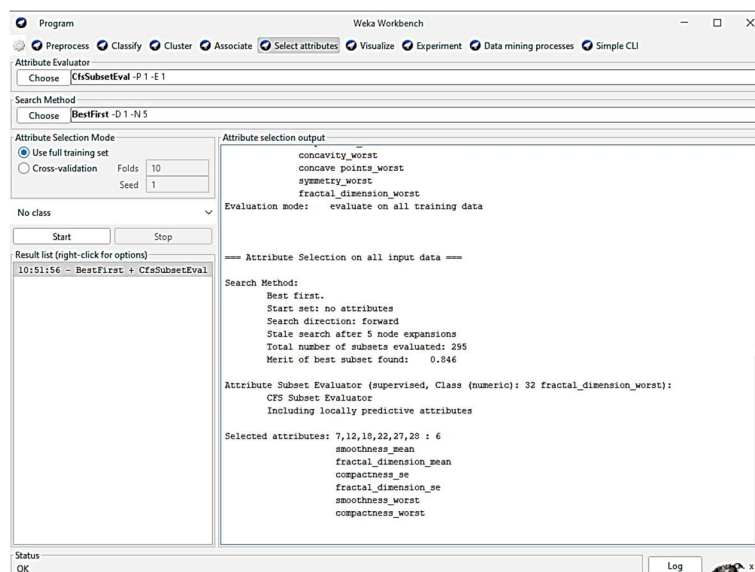


Рис. 4. Визначення оптимальної підмножини

Джерело: розроблено автором.

Ці результати показують, що навіть за умови зменшення кількості вхідних даних до невеликої підмножини можна зберегти високу точність аналізу.

У наступному експерименті для побудови моделі було застосовано RandomForest, проте зменшено кількість вхідних даних атрибутів. Замість 32 атрибутів залишено лише 6 найбільш значущих показників. Такий підхід дозволить перевірити, як відбір ознак вплинув на якість прогнозування.

Результати показали, що точність моделі знизилась порівняно з повним набором даних. Коефіцієнт кореляції становив 0,8141, що вказує на достатньо сильний, але вже слабший зв'язок між фактичними та прогнозованими значеннями. Середня абсолютна похибка зросла до 0,062, а корінь середньоквадратичної похибки до 0,0913, що відображає вищий рівень відхилень. Відносні показники похибки також зросли 51,63 і 57,90 %, що свідчить про втрату частини інформації при скороченні числа ознак (рис. 5).

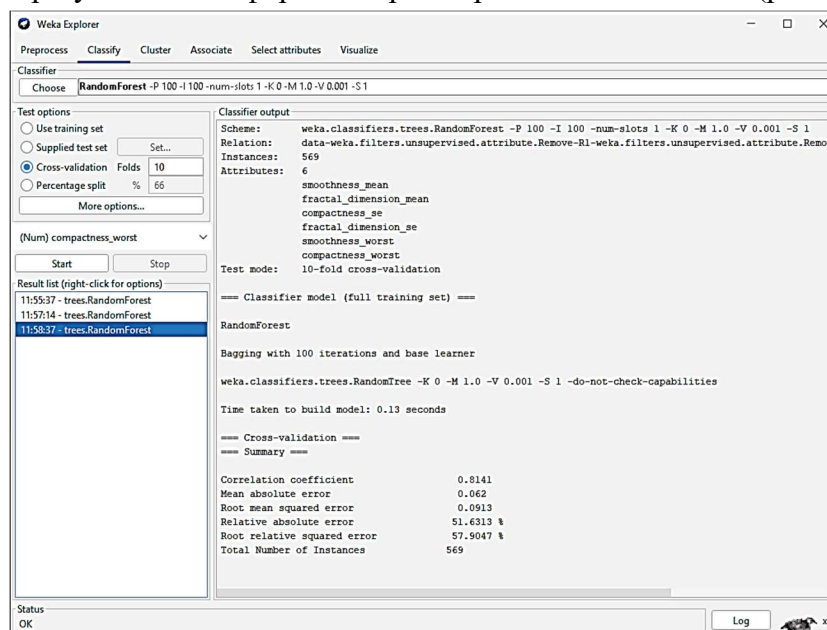


Рис. 5. Побудова моделі з шести атрибутів

Джерело: розроблено автором.

Аналіз трьох класифікацій повного набору даних та скорочених наведено в табл. 3.

Таблиця 3 – Аналіз успішності методів класифікації

| Варіанти моделей    | Correlation coefficient | MAE    | RMSE   | RAE (%) | RRSE (%) |
|---------------------|-------------------------|--------|--------|---------|----------|
| Усі 32 атрибути     | 0,9157                  | 0,0053 | 0,0077 | 39,18   | 42,40    |
| 6 атрибутів з 32    | 0,8141                  | 0,0620 | 0,0913 | 51,90   | 57,90    |
| 31 атрибут (без ID) | 0,9159                  | 0,0052 | 0,0076 | 38,46   | 42,20    |

Джерело: розроблено автором.

**Висновки.** Запропоноване поєднання технологій ПЛІС, IoT та аналізу за допомогою застосунку WEKA має великий потенціал. Такий підхід допоможе поліпшити своєчасне діагностування раку молочної залози на ранній стадії. У ході дослідження було зроблено порівняння чотирьох алгоритмів для стиснення медичних даних, а саме зображень раку молочної залози. Порівняння одного зображення дозволило виявити, що алгоритми JPEG-LS та Deep Learning Lossy продемонстрували однакові результати, коефіцієнт стиснення 1,97. Попри те, що два алгоритми продемонстрували однакові результати, Deep Learning Lossy не є стандартом DICOM. Незважаючи на це, алгоритм Deep Learning Lossy має великий потенціал у застосуванні для медичних даних.

Проведена кластеризація за допомогою застосунку WEKA. Для кластеризації та відбору ключових ознак було взято набір даних, який містить 569 прикладів та 32 атрибути. Було виконано три кластеризації: повного набору даних, набору даних із видаленим ідентифікатором пацієнта та набором, який складається з тих атрибутів, які було відібрано за допомогою CFsSubsetEval у поєднанні з пошуком Best First з початкового набору. Можна зробити висновок, що хоча кластеризація відібраних ознак показала непоганий результат, все одно він вимагає доопрацювання у подальшій праці. Класифікація, яка містила 31 атрибут, продемонструвала коефіцієнт кореляції 0,9159, що є добрим результатом. Також було підтверджено, що атрибут з назвою «ідентифікатор пацієнта» не несе в собі корисної інформації для аналізу.

### Заява про використання генеративного ШІ та технологій на основі ШІ в процесі написання тексту статті

Під час написання цього матеріалу автор використовував ChatGPT для упорядкування думок стосовно постановки проблеми. Після використання цього інструменту/сервісу автор переглянув та відредагував зміст за потреби і взяв на себе повну відповідальність за зміст публікації.

### Список використаних джерел

1. Gayathri, R., Vijaya Madhari, M., & Christy Bobby, T. (2015). FPGA based diagnosis and classification of breast disease using thermographic images. *International Journal of Scientific Research and Management*, 3(6), 3139-3142.
2. Vaithianathan, Muthukumaran & Udkar, Shiv & Reddy, Manjunath & Roy, Deepanjan & Rajasekaran, Senkadir. (2024). FPGA-Based Smart Health Monitoring Systems for Wearable Devices. *International Journal of intelligent systems and applications in engineering*. <https://doi.org/10.13140/RG.2.2.33185.62567>.
3. Alshammari, M., & Mezher, M. (2020). A Comparative Analysis of Data Mining Techniques on Breast Cancer Diagnosis Data using WEKA Toolbox. *International Journal of Advanced Computer Science and Applications*, 11(8). <https://doi.org/10.14569/ijacsa.2020.0110829>.
4. Mhaouch, A., Gtifa, W., & Machhout, M. (2025a). FPGA Hardware Acceleration of AI Models for Real-Time Breast Cancer Classification. *AI*, 6(4), 76. <https://doi.org/10.3390/ai6040076>.
5. Pagliari, D. J., Casu, M. R., & Carloni, L. P. (2017). Accelerators for Breast Cancer Detection. *ACM Transactions on Embedded Computing Systems*, 16(3), 1-25. <https://doi.org/10.1145/2983630>.
6. Majji, R., G., O. P. P., Rajeswari, R., & R., C. (2023). Smart IoT in Breast Cancer Detection Using Optimal Deep Learning. *Journal of Digital Imaging*. <https://doi.org/10.1007/s10278-023-00834-9>.
7. Mhaouch, A., Gtifa, W., & Machhout, M. (2025). FPGA Hardware Acceleration of AI Models for Real-Time Breast Cancer Classification. *AI*, 6(4), 76. <https://doi.org/10.3390/ai6040076>.

8. Wong, E.-H., Zakaria, F. F., Mustapa, M., Warip, M. N. M., & Ehkan, P. (2023). Breast cancer classification using deep learning and FPGA inferencing. In *International Conference on Biomedical Engineering (ICoBE 2021)*. AIP Publishing. <https://doi.org/10.1063/5.0111204>.

### References

1. Gayathri, R., Vijaya Madhari, M., & Christy Bobby, T. (2015). FPGA based diagnosis and classification of breast disease using thermographic images. *International Journal of Scientific Research and Management*, 3(6), 3139-3142.
2. Vaithianathan, Muthukumaran & Udkar, Shiv & Reddy, Manjunath & Roy, Deepanjan & Rajasekaran, Senkadir. (2024). FPGA-Based Smart Health Monitoring Systems for Wearable Devices. *International Journal of intelligent systems and applications in engineering*. <https://doi.org/10.13140/RG.2.2.33185.62567>.
3. Alshammari, M., & Mezher, M. (2020). A Comparative Analysis of Data Mining Techniques on Breast Cancer Diagnosis Data using WEKA Toolbox. *International Journal of Advanced Computer Science and Applications*, 11(8). <https://doi.org/10.14569/ijacsa.2020.0110829>.
4. Mhaouch, A., Gtifa, W., & Machhout, M. (2025a). FPGA Hardware Acceleration of AI Models for Real-Time Breast Cancer Classification. *AI*, 6(4), 76. <https://doi.org/10.3390/ai6040076>.
5. Pagliari, D. J., Casu, M. R., & Carloni, L. P. (2017). Accelerators for Breast Cancer Detection. *ACM Transactions on Embedded Computing Systems*, 16(3), 1-25. <https://doi.org/10.1145/2983630>.
6. Majji, R., G., O. P. P., Rajeswari, R., & R., C. (2023). Smart IoT in Breast Cancer Detection Using Optimal Deep Learning. *Journal of Digital Imaging*. <https://doi.org/10.1007/s10278-023-00834-9>.
7. Mhaouch, A., Gtifa, W., & Machhout, M. (2025). FPGA Hardware Acceleration of AI Models for Real-Time Breast Cancer Classification. *AI*, 6(4), 76. <https://doi.org/10.3390/ai6040076>.
8. Wong, E.-H., Zakaria, F. F., Mustapa, M., Warip, M. N. M., & Ehkan, P. (2023). Breast cancer classification using deep learning and FPGA inferencing. In *International Conference on Biomedical Engineering (ICoBE 2021)*. AIP Publishing. <https://doi.org/10.1063/5.0111204>.

Отримано 19.09.2025

UDC 004.85:004.77

### Honcharov Denys

Phd Student, Department of Computer Engineering  
Petro Mohyla Black Sea National University (Mykolaiv, Ukraine)

E-mail: [honcharov.denys@chmnu.edu.ua](mailto:honcharov.denys@chmnu.edu.ua). ORCID: <https://orcid.org/0009-0004-1200-6677>

## BREAST CANCER DIAGNOSTICS: IOT SYSTEM WITH FPGA AND WEKA ANALYTICS

Breast cancer remains one of the most common diseases among women. Timely diagnosis will help reduce mortality rates and reduce treatment costs, as the longer the disease remains undiagnosed, the cost of treatment increases more. The article proposes an approach to creating a portable device based on the FPGA system. This device can be used at home, without the need to make an appointment with a doctor once more. A comparative analysis of four image compression algorithms was made. When compressing the same image with different algorithms, the best result was achieved using JPEG-LS and Deep Learning Lossy. Compression coefficient scores were 1,97. It should be noted that compression algorithms with large losses will not give the desired result, since there is a chance to lose an important part of the image for a correct diagnosis. The WEKA application was used for data analysis. The first study was conducted on a complete data set of 596 examples and 32 attributes. The analysis was carried out using the RandomForest algorithm with tenfold validation. The results demonstrated high efficiency: the correlation coefficient was 0,9157, the mean absolute error – 0,0053, and the root of the mean square error – 0,0077. The second experiment consisted of 596 instances and 31 attributes (without patient ID). The result obtained was more accurate. The correlation coefficient is 0,9159, the mean absolute error is – 0,0052, and the root of the mean square error is – 0,0076. The third experiment differed from the previous ones. Before starting clustering, a search was performed for the attributes that have the greatest value. Using the CFsSubsetEval method in combination with the BestFirst search, 6 attributes were selected from the initial dataset. These indicators included indicators: smoothness\_mean, fractal\_dimension\_mean, compactness\_se, fractal\_dimension\_se, smoothness\_worst and compactness\_worst. As a result of using a set of 6 attributes, the accuracy of the model decreased, but still remained high enough, which indicates the possibility of maintaining the accuracy at a sufficient level even with a reduced amount of input data.

**Keywords:** FPGA; WEKA; information technology; image compression; jpeg-ls.

Fig.: 5. Table: 3. References: 8.