

УДК 004.93

В.В. СПІРИНЦЕВ, А.Ю. ГОГОЛЮК
Дніпропетровський національний університет ім.Олеся Гончара**АНАЛІЗ МАМОГРАФІЧНИХ ЗНІМКІВ З ВИКОРИСТАННЯМ
АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ**

Здійснено порівняльний аналіз алгоритмів кластеризації (k -means, FCM, HCM, GK, FAFCM, FCM-SOM, алгоритму гібридної кластеризації), досліджено ступінь їх ефективності (при варіюванні різних параметрів), доступності і здатності до практичного використання при аналізі мамографічних знімків.

Ключові слова: кластеризація, мамографія, FCM, HCM, FAFCM, K-MEANS, гібридна кластеризація.

В.В. СПИРИНЦЕВ, А.Ю. ГОГОЛЮК
Днепропетровский национальный университет им.Олеся Гончара**АНАЛИЗ МАММОГРАФИЧЕСКИХ СНИМКОВ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ
КЛАСТЕРИЗАЦИИ**

Осуществлен сравнительный анализ алгоритмов кластеризации (k -means, FCM, HCM, GK, FAFCM, FCM-SOM, алгоритма гибридной кластеризации), исследована степень их эффективности (при варьировании различных параметров), доступности и применимости к практическому использованию при анализе маммографических снимков.

Ключевые слова: кластеризация, маммография, FCM, HCM, FAFCM, K-MEANS, гибридная кластеризация.

V.V. SPIRINTSEV, A.Yu. HOHOLIUK
Dnepropetrovsk National University named after Oles Honchar**THE ANALYSIS OF MAMMOGRAPHIC ROENTGENOGRAM
WITH USING THE CLUSTERING ALGORITHMS**

The comparative analysis of clustering algorithms is carried out (k -means, FCM, HCM, GK, FAFCM, FCM-SOM, an algorithm of a hybrid clustering). The level of their efficiency (for varying parameters), accessibility and the ability of practical use while analyzing the mammographic roentgenograms is studied.

Keywords: clustering, mammography, FCM, HCM, FAFCM, K-MEANS, hybrid clustering.

Постановка проблеми

Одним із найпоширеніших видів онкологічних захворювань серед жінок за даними ВООЗ (Всесвітній організації охорони здоров'я) є рак молочної залози. Для діагностики онкологічних захворювань застосовують мамографію [1] – неінвазивний рентгенологічний метод дослідження молочних залоз. Існує багато видів мамографії, кожний з яких має свої переваги та недоліки, та специфіку діагностування. Аналіз знімків при цьому не дає однозначного діагнозу, оскільки залежить від кваліфікації лікаря та його суб'єктивної думки. Використання алгоритмів кластеризації при аналізі мамографічних зображень дозволяє знизити ризики встановлення хибного діагнозу пацієнтам та усунути суб'єктивну складову людського фактору [2].

Аналіз останніх досліджень та публікацій

Зараз використовується безліч методів кластеризації, що істотно відрізняються між собою за різними показниками: тип вихідних даних, чутливість, швидкість дії, кількість управляючих параметрів та ін. [2]. Серед інструментів, що найбільш часто використовуються для виконання кластеризації, виділяють методи нечіткої логіки, нейронні мережі та нейро-фаззі алгоритми (які включають переваги як методів нечіткої логіки так і нейронних мереж). Слід відзначити, що неоднозначність процедури дефазифікації нечітких методів та нейро-фаззі алгоритмів істотно впливає на ефективність та якість кінцевого результату [3]. В даній роботі пропонується здійснювати аналіз мамографічних зображень на основі алгоритмів кластеризації (k -means, FCM, HCM, GK, FAFCM, FCM-SOM, алгоритму гібридної кластеризації [4]) при різних параметрах (експоненційна вага, вибір метрики при динамічному стисненні) та дослідити ефективність даних алгоритмів для виявлення новоутворень у молочній залозі та мікрокальцинатів (відкладення солей кальцію).

Формулювання мети дослідження

Здійснити порівняльний аналіз результатів роботи алгоритмів кластеризації та виявити оптимальні параметри для виділення новоутворень молочної залози на мамографічних знімках.

Викладення основного матеріалу дослідження

На рис.1 наведено результат роботи алгоритмів (*k*-means, FCM, HCM, GK, FAFCM, FCM-SOM, алгоритму гібридної кластеризації) при однакових параметрах (експоненціальна вага дорівнює 2; кількість нечітких кластерів – 5; критерій зупину – 0,00001) для виявлення новоутворень. Для всіх алгоритмів, крім *k*-means, використано спосіб візуалізації порівняння з початковими даними [5], а для *k*-means – візуалізація на основі кластерів [6] (це обумовлено тим, що в даному алгоритмі не розраховується матриця належності).

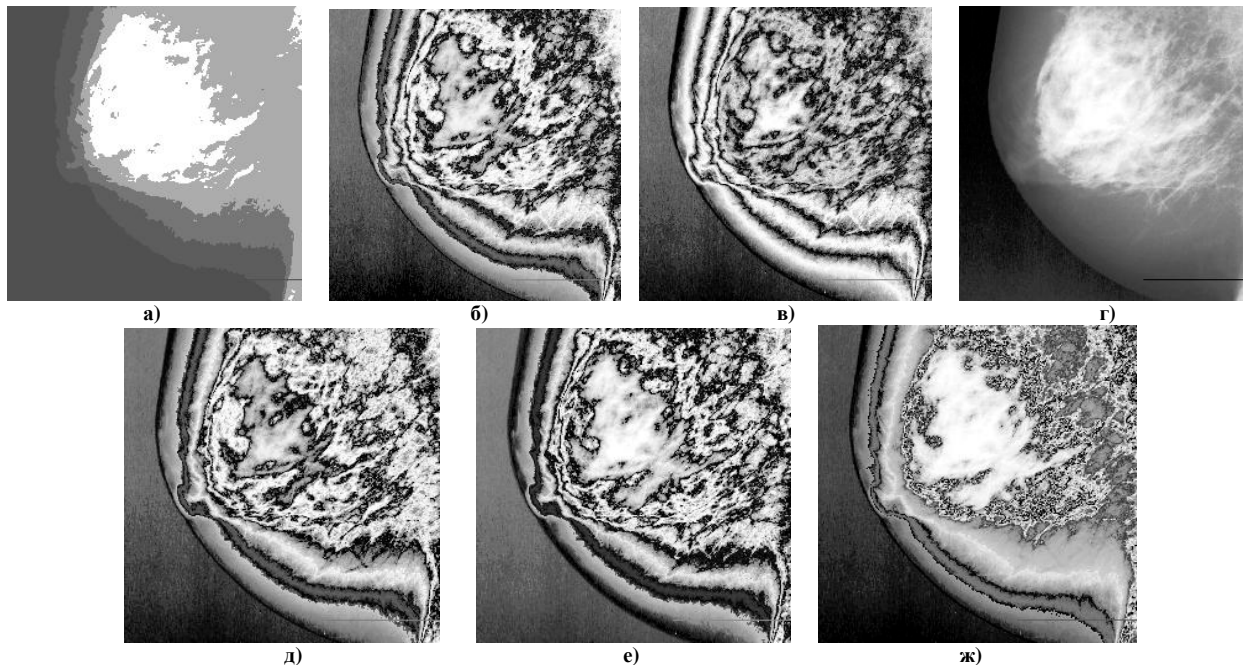


Рис. 1. Результат роботи алгоритмів кластеризації: а) – *k*-means; б) – FCM; в) – HCM; г) – FAFCM; д) – GK; е) – FCM-SOM; ж) – алгоритм гібридної кластеризації

Результат аналізу (див. рис.1) показав неефективність роботи алгоритмів *k*-means та FAFCM у порівнянні з іншими алгоритмами. Щодо алгоритму GK (хоча він і використовується здебільшого для багатовимірних зображень) результат його роботи із напівтоновим зображенням вражає, оскільки необхідна область чітко виділена.

В табл. 1 наведено результати швидкодії запропонованих алгоритмів кластеризації.

Таблиця 1

Результати швидкодії алгоритмів кластеризації	
Алгоритм кластеризації	Час виконання, с
<i>k</i> -means	0.36
FCM	5.8702
HCM	1.424
FAFCM	3.0188
GK	2.1423
FCM-SOM	7.9572
Гібридна кластеризація	6.2245

Аналіз даних з табл. 1 показує, що швидкість дії найвища у алгоритму *k*-means при найгіршому результаті якості отриманого зображення; далі за швидкістю відзначимо алгоритми HCM та GK (в 2-3 рази швидше за стандартний FCM), які також ефективно виділяють область новоутворення молочної залози. Щодо алгоритмів, які мають у своїй архітектурі нейронну мережу, швидкість їх дії є найнижчою, що пояснюється складним механізмом роботи.

Подальші дії було спрямовано на дослідження впливу експоненціальної ваги для базового алгоритму FCM на результат обробки мамографічного зображення (рис.2). Загалом, цей параметр використовують для знаходження матриці відстаней. Було обрано наступні параметри: кількість нечітких кластерів дорівнювала 5; критерій зупину складав 0,00001; експоненціальну вагу змінювали від 2 до 7.

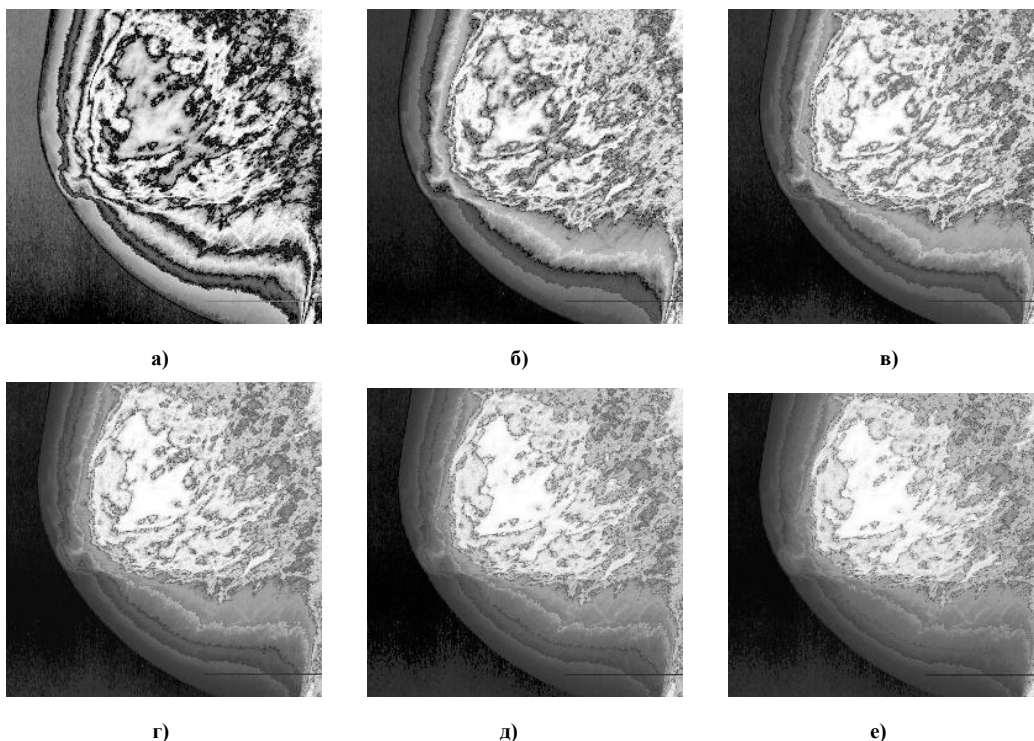


Рис. 2. Візуалізація результатів роботи алгоритму FCM при зміні експоненціальної ваги: а) – 2; б) – 3; в) – 4; г) – 5; д) – 6; е) – 7

Результат аналізу (див. рис.2) показав: збільшення значення експоненціальної ваги знижує якість результуючого зображення (виділення кластерів) та збільшує час обробки. Оптимальним значенням параметра є значення 2 і 3.

Інший напрям досліджень стосувався вивчення впливу метрики при динамічному стисненні [4] на результат обробки (рис.3). Було обрано наступні параметри: кількість нечітких кластерів дорівнювала 20 (із надлишком); критерій зупину складав 0,00001; експоненціальна вага – 2; відсоткове співвідношення між максимальною та мінімальною відстанню центрів нечітких кластерів становило 0,05. Крім Евклідової метрики, також використано зважену Евклідову відстань.

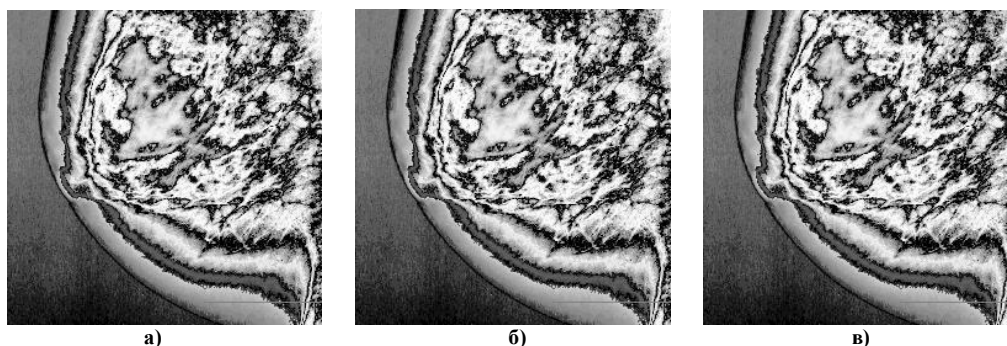


Рис. 3. Візуалізація результату роботи алгоритму FCM при процентному співвідношенні: а) – 10%; б) – 50%; в) – 90%

Таблиця 2

Порівняльний аналіз часу роботи за різних відстаней

Евклідова відстань,с	Зважена Евклідова відстань,с
8.5321	2.9632
8.2014	2.5571
8.3821	2.6457
8.3814	2.5686

Із рис.3 та табл. 2 видно, що використання звичайної Евклідової відстані є більш ефективним у порівнянні зі зваженою Евклідовою відстанню, проте час виконання алгоритму є майже у тричі більшим. Тому постає проблема вибору між якістю результату та часом виконання. Враховуючи специфіку практичної спрямованості (обробка мамографічного знімку), вважаємо за потрібне надати перевагу якості результату обробки.

Також виконано порівняння роботи алгоритму нечіткої FCM та гібридної кластеризації [4] з метою отримання відповіді на питання: чи доцільно використовувати гібридну кластеризацію для виділення новоутворень на мамографічних знімках (рис.4). Для цих двох алгоритмів було обрано наступні параметри: кількість нечітких кластерів дорівнює 20 (із надлишком); критерій зупину 0,00001; експоненційна вага 2; процентне співвідношення між максимальною та мінімальною відстанню центрів нечітких кластерів 0,05. Для нейронної мережі SOM, що входить до алгоритму гібридної кластеризації використано наступні параметри: числове значення, що задає спосіб збільшення кількості кластерів перед використанням SOM дорівнює 2; для методу стиснення кількості кластерів після використання SOM обрано вибір нейронів із максимальною кількістю належних до них початкових даних.

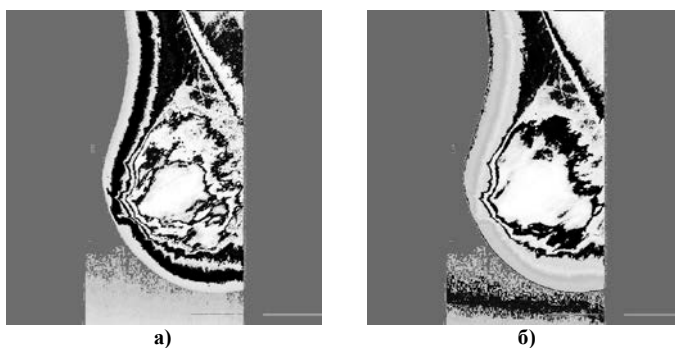


Рис. 4. Візуалізація результатів роботи алгоритмів: а) – FCM; б) – гібридної кластеризації

Із рис. 4 очевидно, що алгоритм гібридної кластеризації не дає змоги дешифрувати мамографічний знімок через високі піки контрасту в області новоутворення (рис. 5).

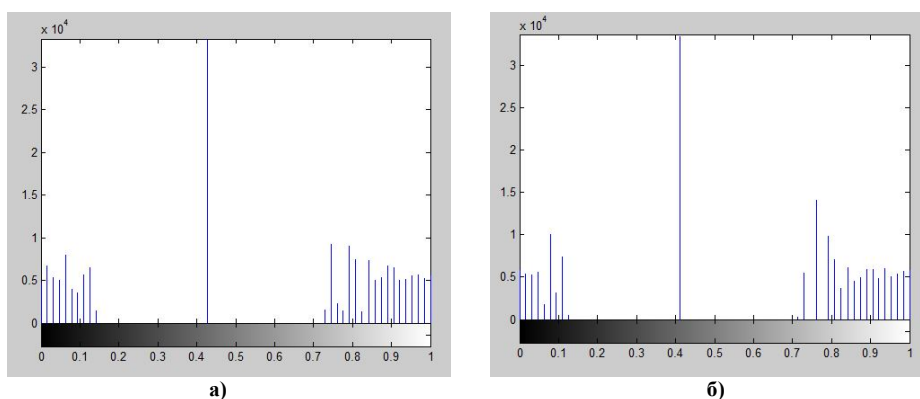


Рис. 5. Гістограми результатів обробки мамографічного знімку алгоритмом: а) – FCM; б) – гібридної кластеризації

Подальші дослідження (рис.6) було спрямовано на застосування алгоритмів кластеризації FCM, HCM, FCM-SOM (при однакових параметрах - попередній дослід, метод візуалізації - порівняння із початковими даними) для виділення мікрокальцинатів (відкладення солей кальцію), що зустрічається досить часто на практиці. Дані відкладення з'являються на місці загиблих чи змінених тканин в наслідок запальних процесів. Якщо вони маленькі (зернисті), то на мамограмі їх важко розпізнати або можна зплутати із дефектом плівки.

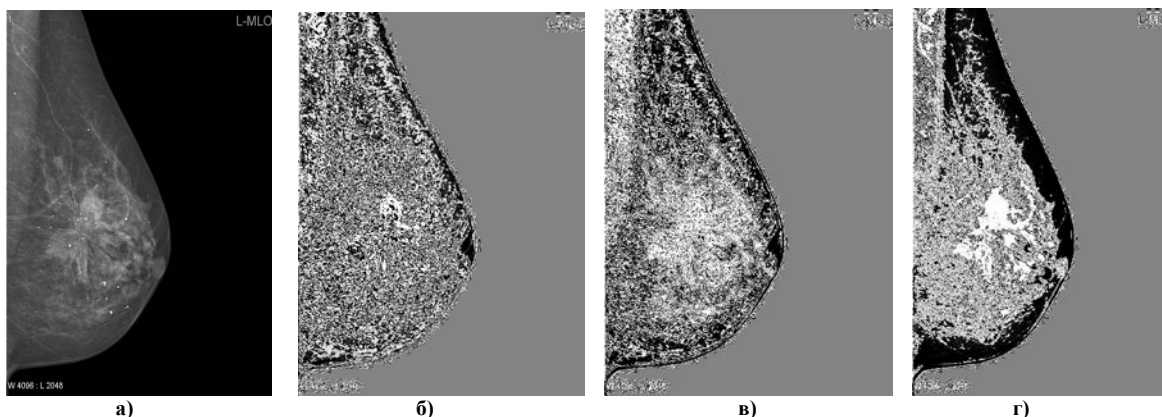


Рис.6. Візуалізація результатів роботи алгоритмів: а) – вхідні дані; б) – HCM; в) – FCM; г) – FCM-SOM

Жодний із наведених вище алгоритмів не дає змоги виділити такі маленькі об'єкти як мікрокальцинати (рис. 6). Отриманий результат більше схожий на шум, хоча контури пухлини чітко виділені. Тому робимо висновок, що алгоритми кластеризації FCM, HCM, FCM-SOM є неефективними при виділенні мікрокальцинатів на мамографічних знімках.

Висновки

В даній роботі здійснено аналіз мамографічних зображень з використанням алгоритмів кластеризації: *k*-means, FCM, HCM, GK, FAFCM, FCM-SOM, алгоритму гібридної кластеризації. Результат аналізу показав неефективність роботи алгоритмів *k*-means та FAFCM. Стосовно швидкодії запропонованих алгоритмів кластеризації: швидкість дії найвища у алгоритма *k*-means, при найгіршому результаті якості отриманого зображення; далі за швидкодією відзначимо алгоритми HCM та GK (в 2-3 рази швидше за стандартний FCM), які також ефективно виділяють область новоутворення молочної залози. Щодо алгоритмів, які мають у своїй архітектурі нейронну мережу – швидкість дії найнижча, що пояснюється складним механізмом роботи. В роботі експериментально показано, що збільшення значення експоненційної ваги для базового алгоритму FCM знижує якість результуючого зображення (виділення кластерів) та збільшує час обробки; визначено оптимальне значення параметру експоненційної ваги. Результат аналізу впливу метрики при динамічному стисненні на результат обробки показав, що використання звичайної Евклідової відстані є більш ефективним у порівнянні зі зваженою Евклідовою відстанню, проте час виконання алгоритму є майже у тричі більшим. Також було порівняно роботу алгоритму нечіткої та гібридної кластеризації та зроблено висновок про їх неефективність при виділенні новоутворень. В роботі експериментально доведено неефективність застосування алгоритмів кластеризації FCM, HCM, FCM-SOM для виділення мікрокальцинатів на мамографічних знімках.

Результати роботи можуть бути використані при розробці нових універсальних або модифікації вже існуючих алгоритмів з визначеними критеріями їх ефективності для обробки конкретного типу зображення.

Список використаної літератури

1. Садыков С.С. Автоматизированная обработка и анализ маммографических снимков: монография / С.С. Садыков, Ю.А. Буланова, Е.А. Захарова. – Владимир: Изд-во ВлГУ, 2014. – 208 с.
2. Гонсалес Р. Цифровая обработка изображений / Р.Гонсалес, Р. Вудс; [пер. с англ. Под ред. П.А. Чочиа]. – М.: Техносфера, 2006. – 1070 с.
3. Леоненков А. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А. Леоненков – С.Пб.: БХВ-Петербург, 2003. – 719 с.
4. Ахметшина Л.Г. Повышение чувствительности алгоритма гибридной нечеткой кластеризации на основе вычисления взвешенного Евклидова расстояния / Л.Г. Ахметшина, А.А. Егоров // Математичне та програмне забезпечення інтелектуальних систем: IX міжнародна науково-практична конференція (м. Дніпропетровськ, 16-18 листопада 2011 р.). – Дніпропетровськ, 2011. – С. 13–14.
5. Егоров А.О. Візуалізація результатів нечіткої кластеризації на основі порівняння з початковими даними / А.О. Егоров // Математичне та програмне забезпечення інтелектуальних систем: VIII міжнародна науково-практична конференція (м. Дніпропетровськ, 12-14 листопада 2008 р.) – Дніпропетровськ, 2008. – С. 117–118.
6. Лепский А.Е. Математические методы распознавания образов / А.Е. Лепский, А.Г. Броневиц. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – 155 с.