

Подані вище результати вкотре показують, що утворені внаслідок опромінення радіаційні дефекти в досліджуваних монокристалах CdSb відіграють важливу роль у формуванні їх електричних властивостей. Вивчення цих особливостей створює передумови для врахування описаних ефектів при конструюванні різних напівпровідникових приладів, а також забезпечить реальні способи мінімізації проявів цих ефектів там, де вони небажані.

#### Джерела та література

1. Семенюк А. К. Радіаційні ефекти в багатодолинних напівпровідниках / А. К. Семенюк. – Луцьк : Надстир'я, 2001. – 323 с.
2. Полупроводниковые соединения  $A^2B^5$  / В. Б. Лазарев, В. Я. Шевченко, Я. Х. Гринберг, В. В. Соболев. – М. : Наука, 1978. – 256 с.
3. Влияние адсорбции на электрические свойства CdSb / Т. С. Гертович, В. Г. Никулина, А. Ф. Семизоров и др. // Физическая электроника. – 1973. – № 6. – С. 52–58.
4. Влияние радиационных дефектов на некоторые электрические и оптические свойства монокристаллов антимида кадмия / С. А. Федосов, Г. Е. Давидюк, В. В. Божко и др. // Неорганические материалы. – 1996. – Т. 32, № 11. – С. 1333–1337.
5. Зарубін Л. І. Електричні властивості CdSb, легованого Au при низьких температурах / Л. І. Зарубін, О. Г. Міселюк // Укр. фіз. журн. – 1967. – Т. 12, № 4. – С. 607–612.
6. Емцев В. В. О возможном механизме рассеяния носителей заряда в германии с точечными дефектами / В. В. Емцев, М. И. Клигер, Т. В. Машовец // Письма в ЖЭТФ. – 1974. – Т. 19. – С. 575–579.
7. Баранський П. І. Фізичні властивості кристалів кремнію та германію в полях ефективного зовнішнього впливу / П. І. Баранський, А. В. Федосов, Г. П. Гайдар. – Луцьк : Надстир'я, 2000. – 280 с.

**Коваль Юрий. Влияние дефектов радиационного происхождения на электрические свойства монокристаллов антимида кадмия.** Исследовано влияние  $\gamma$ -облучения на кинетические параметры монокристаллов антимида кадмия. Измерялись удельная проводимость и эффект Холла, что позволило получить зависимости постоянной Холла, концентрации носителей заряда и их подвижности от температуры. Обнаружено факт резкого увеличения значений подвижности носителей заряда в облученных кристаллах при дозах облучения до  $4 \cdot 10^{18} \gamma$  кв/см<sup>2</sup>, которое объясняется наличием так называемого «эффекта малых доз». Отмечено, что увеличение подвижности носителей заряда, наблюдаемое в данной работе, возникает не за счет повышения упорядоченности кристаллов при их радиационной обработке, а за счет снижения эффективности рассеяния носителей заряда на ионных остатках примеси при частичной нейтрализации заряда противоположным по знаку зарядом дефектов.

**Ключевые слова:** антимида кадмия,  $\gamma$ -облучение, кинетические параметры.

**Koval Yurii. The Influence of Radiation-induced Defects on the Electrical Properties of Cadmium Antimony Monocrystals.** Influence of  $\gamma$ -irradiation on the kinetic parameters of cadmium antimony monocrystals has been explored. Conductivity and Hall effect have been measured, that allowed to get temperature dependences of Hall constant, concentration of charge transmitters and their mobility. The sharp increase of transmitters mobility in the crystals, exposed to the  $\gamma$ -irradiation at the doses up to  $4 \cdot 10^{18} \gamma$ -quantum/cm<sup>2</sup>, has been discovered. We explain this fact by the presence of the so-called «effect of small doses». It is marked, that the increase of charge transmitters mobility, observed in this work, arises up not due to the increase of ordering of crystals after their radiation treatment, but due to the decline of efficiency of dispersion of charge transmitters on ionic residues of admixture during partial neutralization of charge by the opposite charge of defects.

**Key words:** cadmium antimony,  $\gamma$ -irradiation, kinetic parameters.

Стаття надійшла до редколегії  
02.10.2014 р.

УДК 519.711:616.12

Дмитрий Кунашев

### Разработка метода автоматической классификации клеток

Статья посвящена решению одной из главных проблем диагностики и классификации (распределения) здоровых и раковых клеток. Материалом исследования выступили 1346 изображений клеток молочной железы

© Кунашев Д., 2014

без выраженных патологических изменений (данные 12 обследуемых), а также 1944 раковых клеток (данные 20 обследуемых). На основании проведенных исследований представлены результаты работы над методом автоматической классификации раковых клеток с помощью классификационной модели. Проведен анализ различий в изображениях, полученных в результате выделения ядра у больных и здоровых клеток. На основании простых статистических параметров произведен расчет классификационных моделей и оценена их эффективность. Доказано, что данная модель с точностью 76,1 % разделяет все изображения клеток на больные и здоровые, а поэтому с успехом может быть использована в качестве метода автоматической классификации изображений клеток больных и здоровых обследуемых.

**Ключевые слова:** раковые клетки, классификация, дискриминантный анализ, логистическая регрессия.

**Постановка научной проблемы и её значение.** При диагностике раковых заболеваний возникает необходимость в автоматизированном способе классификации изображений клеток на здоровые и больные. Очевидна потребность в простом методе, который способен с высокой точностью диагностировать раковое заболевание без вмешательства человека. Данный способ, имея выборку изображений клеток пациента, должен автоматически диагностировать наличие или отсутствие ракового заболевания или опухоли.

**Анализ исследований этой проблемы.** Известные методы решения задачи, классификации изображения имеют большое разнообразие в способах и стратегиях реализации этапов.

Проблема распознавания изображения имеет более чем 20-летнюю историю. Первые работы относятся к 80-м годам XX в. [3]. Однако наиболее существенное развитие это направление получило в последние несколько лет. Многие из выполненных исследований посвящены развитию стратегий последовательного уточнения запроса и оптимизаций поисковых процедур для изображений в конкретных базах данных большого объема [1–3]. Общая стратегия в разработке таких систем семантического поиска состоит в создании нового запроса, который оптимизируется в процессе диалога с пользователем. Однако для эффективного использования эта стратегия требует сложной трансформации базы данных в модель с взвешенным набором весов для терминологических переменных.

Альтернативный подход не супервизорного обучения, основанный на использовании самоорганизующихся карт Кохонена, рассматривается в работах [4; 5]. Достоинство этого метода в том, что в процессе обучения системы не требуется взаимодействия с человеком-оператором. Но вследствие ограниченности правил формирования карт это, с другой стороны, приводит к снижению разделяющей мощности системы распознавания. Кроме того, для восприятия результата классификации человеком необходимо выполнить трактовку автоматически вырабатываемых семантических понятий, что не всегда возможно.

**Цель статьи.** В связи с вышеизложенным, целью нашей работы является разработка простого метода автоматической классификации клеток (разделение на здоровые и раковые), не требующего больших вычислительных мощностей и участия человека-оператора для работы с системой.

**Изложение основного материала и обоснование полученных результатов исследования. Материалы и методы.** Материалом исследования выступили 1346 изображений клеток молочной железы без выраженных патологических изменений (данные 12 обследуемых), а также 1944 раковых клеток (данные 20 обследуемых). Изображения были поданы в разных цветовых фильтрациях и с разными показателями контрастности.

Предварительная обработка изображений заключалась в выделении ядра клетки и переводе всех изображений в единый цветовой фильтр.

Выделение ядра клетки осуществлялось методом последовательного сканирования полутоновых изображений [5–7] (рис. 1). Данный метод имеет неплохое быстродействие, хорошо настраивается под входные данные, т. е. является устойчивым к разрывам границ, рассчитан на нахождение кластеров с различным геометрическим смыслом, и имеет невысокий уровень человеческого вмешательства.

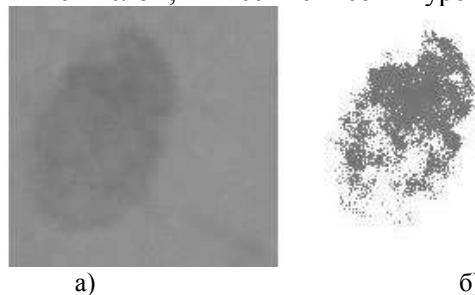


Рис. 1. Результаты обработки изображения по выделению ядра: а) до обработки, б) после обработки

Результатом обработки изображений и их преобразования были текстовые документы, данные в которых представляли собой матрицу  $160 \times 160$ , где каждому пикселю изображения сопоставлялось численное выражение интенсивности серого. Пиксели, находящиеся вне выделенного ядра, из рассмотрения исключались.

Анализ данных для построения моделей был произведен в два этапа:

- 1) проверка различий между группами изображений здоровых и больных обследуемых и выделения параметров для построения классификационной модели;
- 2) построение моделей и проверка их эффективности на тестовой выборке.

Для статистического анализа отличия между клетками, а также для построения классификационной модели был использован статистический пакет обработки данных IBM SPSS Statistics 20.0.

Оценка качества моделей проводилась с помощью кросс-проверки (каждое наблюдение классифицируется функциями, выведенными по всем наблюдениям, за исключением его самого). Рассчитано параметры чувствительности (доля позитивных результатов теста в группе больных пациентов), специфичности (доля негативных результатов теста в группе здоровых пациентов), а также общий процент правильно предсказанных значений.

*Выделение параметров для создания модели классификации.* Исходными данными при разработке классификационной модели служили данные об интенсивности серого в каждом пикселе отдельных изображений. Поскольку качество модели, полученной при использовании одной независимой переменной, крайне низка, это вызвало необходимость дополнительно изучить выборки с целью выделения параметров, которые могут повысить эффективность классификации.

Проверка на нормальность распределения данных об интенсивности серого была проведена тестом Колмогорова–Смирнова ( $p < 0,001$ ). Для сравнения выборок использовались непараметрические критерии различия.

Для сравнения распределений интенсивности серого двух выборок использовался серийный критерий Вальда–Вольфовица. Данный тест используется для проверки нулевой гипотезы о равенстве целого ряда параметров двух выборок, включая медианы и коэффициенты асимметрии. Эмпирические распределения групп представлены на рис. 2.

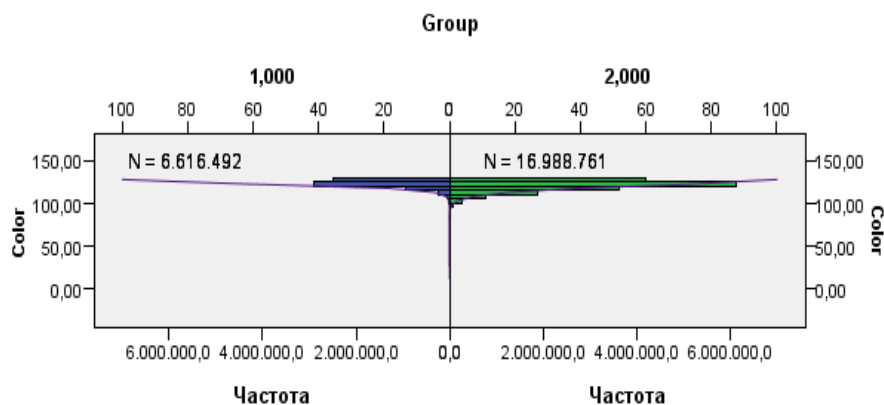


Рис. 2. Распределения интенсивности серого в выборках

Результат сравнения показал значимую разницу в распределении показателя интенсивности серого в группах здоровых и раковых клеток ( $p < 0,001$ ).

Дополнительно было принято решение проверить различия данных выборок критериями разности средних. Применялся медианный критерий для независимых выборок, а также критерий Манна–Уитни (ранговый критерий для двух независимых выборок).

На рис. 3 представлены ящичковые диаграммы, которые характеризуют обе выборки. Результат сравнения показал значимую разницу в данных группах ( $p < 0,001$ ).

В результате данных исследований было доказано, что результаты обработки исходных изображений методом последовательного сканирования полутоновых изображений для получения интенсивности серого у здоровых и раковых клеток отличаются по основным статистическим характеристикам.

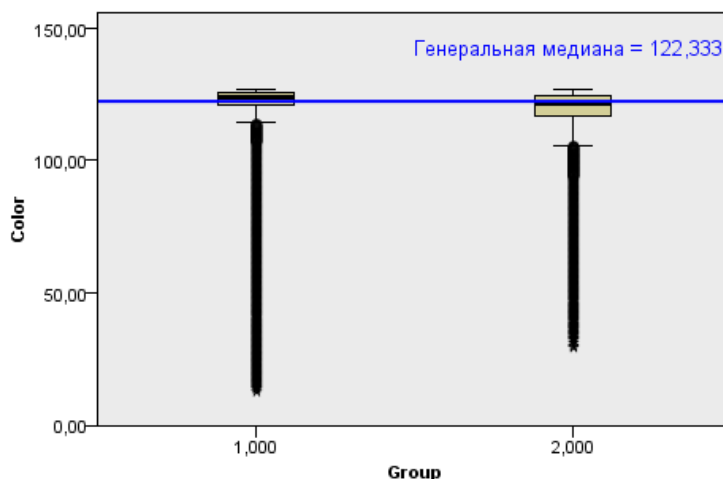


Рис. 3. Ящичковые диаграммы разброса данных в выборках

Для каждой клетки двух выборок были рассчитаны параметры расположения (среднее значения (M), моды (most often – Mo), медианы (middle value – Mv), квантили (25Q, 50Q, 75Q)), разброса (дисперсия (D), стандартное отклонение (StD), минимум (Min), максимум (Max), стандартная ошибка среднего (SeM) и распределения (асимметрия (As) и эксцесс (Exс)). Эти данные (табл. 1) в дальнейшем были использованы для построения классификационной модели.

Таблица 1

Данные для каждой клетки двух выборок

Параметр	Группа		
	здоровые	раковые	
Среднее	122,78	120,10	
Стандартная ошибка среднего	0,00155	0,00145	
Медиана	123,67	121,67	
Мода	127,00	127,00	
Стандартное отклонение	3,99	5,96	
Дисперсия	15,94	35,50	
Асимметрия	-2,82	-1,26	
Стандартная ошибка асимметрии	0,00	0,00	
Эксцесс	31,60	2,23	
Стандартная ошибка эксцесса	0,00	0,00	
Минимум	12,67	29,33	
Максимум	127,00	127,00	
Процентили	25	121,00	117,00
	50	123,67	121,67
	75	125,67	124,67

*Построение классификационной модели.* Для построения классифицирующей функции использовался дискриминантный анализ [9] и логистическая регрессия [10]. Эффективность моделей оценена с помощью показателей чувствительности и специфичности, рассчитанные на тестовой выборке, которая не принимала участия в построении моделей.

Логистическая регрессия – это статистическая модель, используемая для предсказания вероятности возникновения некоторого события путём подгонки данных к логистической кривой. Целью применения данного метода при разработке метода автоматической классификации клеток было получение модели для расчета вероятности принадлежности клетки к группе здоровых или больных. Использовалась бинарная логистическая регрессия с пошаговым включением, что позволило отобрать показатели, которые наиболее точно классифицируют обработанные изображения. Отбор был проведен за 9 шагов, отобрано 7 показателей.

Модель для классификации изображений представлена ниже (1):

$$p = \frac{1}{1 + e^{-y}},$$

где:

$$y = -M \cdot 2,139 - SeM \cdot 9,047 + Mv \cdot 0,380 - As \cdot 0,545 + Exc \cdot 1,785 + Min \cdot 0,070 + 25Q \cdot 0,477 + 150,579. \quad (1)$$

Значимость всех коэффициентов уравнения регрессии меньше 0,05.

Результат анализа (табл. 2) классификационной модели мерами определенности (Кокса & Шела и Наделькеркса), указывающими на ту часть дисперсии, которую можно объяснить с помощью логистической регрессии, составляет 56,3 %.

Таблица 2

Результат анализа (мерами определенности)

Группа		Классификация		Процент корректных, %
		здоровые	больные	
Исходные	Здоровые	1035,0	311,0	76,9
	Больные	447,0	1383,0	75,6
	Здоровые	76,9	23,1	–
	Больные	24,4	75,6	–
Кросс-проверенные	Здоровые	1029,0	317,0	76,4
	Больные	450,0	1380,0	75,4
	Здоровые	76,4	23,6	–
	Больные	24,6	75,4	–

76,1 % исходных сгруппированных наблюдений классифицировано правильно. 75,9 % перекрестно-проверяемых сгруппированных наблюдений классифицировано правильно.

Чувствительность и специфичность на исходной выборке равна 76,9 % и 75,6 % соответственно, при кросс-проверке – 76,4 % и 75,4 %.

Дискриминантный анализ является разделом многомерного статистического анализа, который позволяет изучать различия между двумя и более группами объектов по нескольким переменным одновременно. Он включает в себя методы интерпретации межгрупповых различий – дискриминации, и методы классификации наблюдений по группам.

Построение модели путем применения дискриминантного анализа, так же как и при применении логистической регрессии, производилось методом пошагового включения переменных. Всего было произведено 10 шагов. В окончательную модель было включено 8 переменных (2).

$$d = -M \cdot 0,514 - SeM \cdot 2,167 + SD \cdot 1,071 - D \cdot 0,053 + As \cdot 0,578 + Exc \cdot 0,038 + Min \cdot 0,028 + 25Q \cdot 0,286 + 23,216. \quad (2)$$

Мерой удачности разделения служит корреляционный коэффициент между рассчитанными значениями дискриминантной функции и показателем принадлежности к группе, который для данной модели равен 0,686, что говорит о достаточно удовлетворительной корреляции.

Тест Лямбда Уилкса (выявление отличия друг от друга средних значений дискриминантной функции) показывает результат в  $p < 0,001$ , что свидетельствует об очень значимом различии (табл. 3).

Таблица 3

Результат анализа (Тест Лямбда Уилкса)

Группа		Классификация		Процент корректных, %
		здоровые	больные	
Исходные	Здоровые	991,0	355,0	73,6
	Больные	409,0	1421,0	77,6
	Здоровые	73,6	26,4	–
	Больные	22,3	77,7	–
Кросс-проверенные	Здоровые	988,0	358,0	73,4
	Больные	411,0	1419,0	77,5
	Здоровые	73,4	26,6	–
	Больные	22,5	77,5	–

75,9 % исходных сгруппированных наблюдений классифицировано правильно. 75,8 % перекрестно-проверяемых сгруппированных наблюдений классифицировано правильно.

Чувствительность и специфичность на исходной выборке равна 73,6 % и 77,8 % соответственно, при кросс-проверке – 73,4 % и 77,5 %.

Сравнение результатов моделирования классификационной модели путем бинарной логистической регрессии и дискриминантного анализа показало большую эффективность первой модели. Она может быть использована для автоматического разделения на здоровые и раковые клетки.

**Выводы и перспективы последующего исследования.** В результате анализа исходных изображений здоровых и раковых клеток был получен ряд показателей, которые характеризуют различия в интенсивности серого в двух группах. Они были использованы для построения классификационной модели с помощью бинарной логистической регрессии. Даная модель с точностью 76,1 % разделяет все изображения, что представляет собой высокий результат точности.

В связи с простотой описанного метода, не требующего больших вычислительных мощностей и участия человека-оператора для работы с системой, он может быть использован в качестве метода автоматической классификации изображений клеток больных и здоровых обследуемых.

#### *Источники и литература*

1. Salton G. Introduction to Modern Information Retrieval / G. Salton, M. J. McGill. – New York : McGraw-Hill, 1983. – 458 p.
2. Content-based image retrieval with relevance feedback in MARS / Y. Rui, T. S. Hyang, S. Mehrota // IEEE Int. Conf. Image Processing. – Santa Barbara, CA, 1997. – P. 815–818.
3. Celentano A. Feature integration and relevance feedback analysis in image similarity evaluation / A. Celentano, E. D. Sciasico // Electron Imaging. – 1998. – Vol. 7. – No. 2. – P. 308–317.
4. Gevers T. PieToSeek: Combining color and shape invariant features for image retrieval / T. Gevers, A. W. M. Smeulders // IEEE Trans. Image Processing. – 2000. – Vol. 9. – P. 102–119.
5. Muneesawang P. Automatic Machine Interactions For Content-based Image Retrieval Using a Self-Organizing Tree Map Architecture / P. Muneesawang, L. Guan // IEEE Trans. On Neural Networks. – 2002. – Vol. 13. – No. 4. – P. 93–98.
6. Kohonen T. Self-Organizing Maps (Third Extended Edition) / T. Kohonen. – New York, 2001. – 501 p.
7. Вежневцев А. Выделение связанных областей в цветных и полутоновых изображениях / А. Вежневцев // Компьютерная графика и мультимедиа. – 2003. – Вып. № 1 (5). – С. 58–67.
8. Brunner D. Iterative area filtering of multichannel images / D. Brunner, P. Soille // Image and Vision Computing. – 2007. – Vol. 25, No. 8. – P. 1352–1364.
9. Fu K. S. A survey on image segmentation / K. S. Fu, J. K. Mu // Pattern Recognition. – 1981. – Vol. 13. – No. 1. – P. 3–16.
10. Боровиков В. П. Искусство анализа данных / В. П. Боровиков. – СПб. : Питер, 2005. – 2-е изд. – 525 с.
11. Hosmer D. W. Applied Logistic Regression. – 2nd ed. / D. W. Hosmer, S. Lemeshow. – New York : Chichester, Wiley, 2002. – 392 p.

**Кунашев Дмитро. Розробка методу автоматичної класифікації клітин.** Вирішено одну з головних проблем діагностики та класифікації здорових і ракових клітин. Представлено метод автоматичного розподілу ракових клітин за допомогою класифікаційної моделі. Проаналізовано відмінності у зображеннях, отриманих в результаті виділення ядра у хворих і здорових клітин. На підставі простих статистичних параметрів розраховано класифікаційні моделі та оцінено їхню ефективність. Представлена модель із точністю 76,1 % розподіляє всі зображення клітин, а тому її можна використовувати як метод автоматичної класифікації зображень здорових і хворих клітин.

**Ключові слова:** ракові клітини, класифікація, дискримінантний аналіз, логістична регресія.

**Kunashev Dmytro. Development of a Method the Automatic Classification of Cells.** The article is devoted to one of the main problems of diagnostics of cancer diseases – the search of simple, but at the same time high-accuracy, automated way of classification and distribution of cells on the healthy and ill ones. Research material were 1346 images mammary cells without marked pathological changes (data from 12 subjects) and 1944 cancer cells (data of 20 subjects). The results of work over the method of automated classification of cancer cells with the help of classification model are represented. The analysis of the differences in images, received as a result of extraction of the nucleus at the healthy and ill cells, was conducted. Based upon the simple statistical parameters, calculation of classification models was executed and their effectiveness was evaluated. This model with accuracy 76,1 % divides all images and can be used as a method of automated image classification of cells of sick and healthy examined subjects.

**Key words:** cancer cells, classification, discriminant analysis, logistic regression.

Статья сдана в редколлегию  
19.06.2014 г.