ISSN 2707-1898 (print)
Український журнал інформаційних технологій
Ukrainian Journal of Information Technology
https://science.lpnu.ua/uk/ujit
https://doi.org/10.23939/ujit2021.03.073
Article received 26.04.2021 р.
Article accepted 01.06.2021 р.
UDK 004.93

Correspondence author
O. M. Berezsky
olber62@gmail.com

О. М. Березький[1,2], О. Й. Піцун[2], Г. М. Мельник[2], Т. В. Дацко[3]

[1] *Національний університет "Львівська політехніка", Львів, Україна*
[2] *Західноукраїнський національний університет, Тернопіль, Україна,*
[3] *Тернопільський національний медичний університет, Тернопіль, Україна*

ЗАСТОСУВАННЯ МЕТОДУ ЛІНІЙНОЇ РЕГРЕСІЇ ДЛЯ АНАЛІЗУ КІЛЬКІСНИХ ХАРАКТЕРИСТИК ЦИТОЛОГІЧНИХ ЗОБРАЖЕНЬ

Проаналізовано патологічні стани молочної залози на підставі дослідження цитологічних зображень. Цитологічні зображення – це окремий клас біомедичних зображень і використовуються для діагностування онкологічних захворювань. Для діагностування передракових і ракових станів та тактики лікування лікарі-діагности використовують цитологічні, гістологічні та імуногістохімічні зображення. Для автоматизації процесу діагностування в онкології використовують системи автоматизованої мікроскопії. У системах автоматизованої мікроскопії використовують алгоритми комп'ютерного зору. Останнім часом для класифікації зображень використовують алгоритми машинного навчання Опрацювання мікроскопічних зображень є складним і трудомістким процесом, оскільки зображення характеризуються високим рівнем шуму та відсутністю чітких контурів ядер клітин. Для обчислення кількісних характеристик ядер клітин цитологічних зображень розроблено метод обчислення кількісних характеристик ядер клітин на підставі алгоритмів фільтрації зображень та їх автоматичної сегментації. Для сегментації ядер клітин розроблено архітектуру згорткової нейронної мережі типу U-Net. У цій статті розроблено метод опрацювання цитологічних зображень. Метод складається з шести етапів. На першому етапі відбувається завантаження зображення у пам'ять комп'ютера, на другому – попереднє оброблення зображень. На третьому етапі здійснюється автоматична сегментація зображень на підставі згорткової нейронної мережі типу U-Net. На четвертому етапі виконується обчислення кількісних характеристик ядер клітин. На п'ятому етапі кількісні характеристик ядер клітин зберігаються в базі даних. На шостому етапі застосовують алгоритми лінійної регресії для отримання кількісних характеристик ядер клітин.

Сьогодні лінійна регресія є одним із поширених підходів машинного навчання до аналізу даних. Здійснено порівняльний аналіз застосування кількісних характеристик ядер клітин на підставі лінійної регресії. Наукова новизна роботи полягає у розробленні методу обчислення кількісних характеристик ядер клітин, який містить етапи фільтрації зображень та автоматичної сегментації на підставі використання нейронної мережі типу U-Net. Практичне значення роботи – програмна реалізація модулів попереднього оброблення цитологічного зображення та лінійної регресії. Досліджено, що у сукупності параметрів "площа, довжина головної осі" в 1,4 разу менша похибка RMSE порівняно із сукупністю "площа, периметр".

Ключові слова: автоматична сегментація; цитологічні зображення; нейронна мережа типу U-Net; попереднє оброблення; лінійна регресія.

Вступ

Сучасні технології штучного інтелекту дають змогу автоматизувати процес діагностування в медицині та підвищити його точність. Для діагностування онкологічних хвороб використовують цитологічні зображення, які є окремим класом біомедичних зображень. Цитологічні зображення дають змогу виявляти патологічні стани органів людини, наприклад, молочної залози. Здебільшого для аналізу великого набору цитологічних зображень використовують кількісні характеристики мікрооб'єктів – клітин та ядер клітин. Параметрами ядер клітин є площа, периметр ядра, довжина головної осі, окружність тощо.

Сегментація біомедичних зображень (цитологічних, гістологічних, імуногістохімічних) є важливим етапом опрацювання зображень, який впливає на точність класифікації. Зокрема, якість сегментації істотно впливає на точність обчислення кількісних характеристик ядер клітин. Потреба у використанні сучасних методів сегментації біомедичних зображень зумовлена відсутністю чітких контурів ядер клітин. Складність опрацювання біомедичних зображень передбачає використання сучасних алгоритмів машинного навчання. Тому, незважаючи на вагомі доробки науковців, точність класифікації цього класу зображень є все ще недостатньою.

Розроблення методів аналізу кількісних характеристик ядер клітин – актуальне завдання, оскільки дає змогу пришвидшити постановку діагнозу. Обчислення параметрів кількісних характеристик мікрооб'єктів на цитологічному зображенні є складним і нетривіальним процесом. Тому використання такого сучасного підходу до машинного навчання, як лінійна регресія, дає змогу прогнозувати кількісні характеристики мікрооб'єктів і є *актуальним завданням*.

Об'єкт дослідження – аналіз цитологічного зображення.

Предмет дослідження – методи лінійної регресії для прогнозування кількісних характеристик цитологічних зображень, що дає змогу пришвидшити процес постановки діагнозу.

Мета роботи – розроблення методу та виконання порівняльного аналізу комбінацій параметрів кількісних характеристик ядер клітин на цитологічному зображенні на підставі алгоритмів лінійної регресії.

Для досягнення поставленої мети визначено такі основні *завдання дослідження*:

- проаналізувати диспластичні процеси молочної залози;
- розробити метод обчислення кількісних характеристик ядер клітин на зображенні;
- реалізувати модель нейронної мережі на підставі технології U-Net;
- здійснити порівняльний аналіз результатів лінійної регресії щодо різних комбінацій кількісних показників ядер клітин.

Наукова новизна отриманих результатів дослідження – вперше розроблено метод обчислення кількісних характеристик ядер клітин, який містить етапи фільтрації зображень та автоматичної сегментації на підставі використання нейронної мережі типу U-Net.

Практична значущість результатів дослідження – програмна реалізація модулів попереднього оброблення та лінійної регресії, що дає змогу пришвидшити процес постановки діагнозу.

Аналіз останніх досліджень та публікацій. Першим етапом після отримання зображення є його оброблення та виділення необхідних мікрооб'єктів. Оскільки зображення, одержані за допомогою мікроскопа, характеризуються шумами та нечіткими контурами ядер клітин [3], то необхідно розробити метод оброблення зображень та обчислення кількісних характеристик мікрооб'єктів. Кінцевою метою етапу попереднього оброблення зображень є зменшення імпульсних шумів, вирівнювання гістограми. Chunwei Tian у роботі [11] наводить аналіз шумів на зображеннях та розглядає підходи до їх зменшення за допомогою фільтрації. Аналіз застосування медіанного фільтра, зокрема в медицині, міститься у роботі [9], [10], [13]. Етап сегментації призначений для перетворення вхідного зображення у форматі RGB на бінарне зображення, яке містить білий фон та чорні мікрооб'єкти. Такий формат зображення дає змогу обчислювати кількісні характеристики ядер клітин. У роботах [7], [12] наведено підходи до класифікації медичних зображень, зокрема розглянуто такі алгоритми: Support Vector Machine, K-Nearest Neighbors, Random Forest, Convolutional Neural Network.

Застосування моделей лінійної регресії для опрацювання результатів дослідження описано у роботах [1], [8]. Також автор наводить короткий опис процесу формування вхідного зображення для подальшого опрацювання. Naomi Altman у роботі [2] детально аналізує алгоритми лінійної регресії та за допомогою візуальних інструментів подає результати роботи лінійної регресії.

У роботі [6] проаналізовано процес відбору характеристик цитологічного зображення для діагностування ракових захворювань. У роботах [4], [5] розроблена інтелектуальна система діагностування патологічних станів молочної залози.

Результати дослідження та їх обговорення

Цитологічне дослідження диспластичних процесів молочної залози. Цитологічна діагностика епітеліальних структур допомагає визначити ступінь проліферації епітелію [3]. Виявлення у цитограмі ущільненого апокринового епітелію, сосочкоподібних розростань з проявами формування порожнини або ознак різкої секреторної функції клітин дає змогу цитологічно диференціювати кістозну мастопатію і внутрішньопротокову папілому від фіброаденоми та аденоми. Систематизація цитологічних зображень для мастопатії та фіброаденом показує можливості цитологічного методу щодо формування висновку (діагнозу).

У цитологічних препаратах виявляють епітеліальні клітини порівняно невеликого розміру, за будовою подібні до клітин епітелію ацинусів та протоків незміненої молочної залози. Проте основними ознаками, характерними для долькової непроліферативної мастопатії, визначено наявність сплощеного апокринового епітелію з округлими гіперхромними ядрами, ознаками секреторної активності в клітинах та формуванням сосочкоподібних структур (рис. 1, *а*).

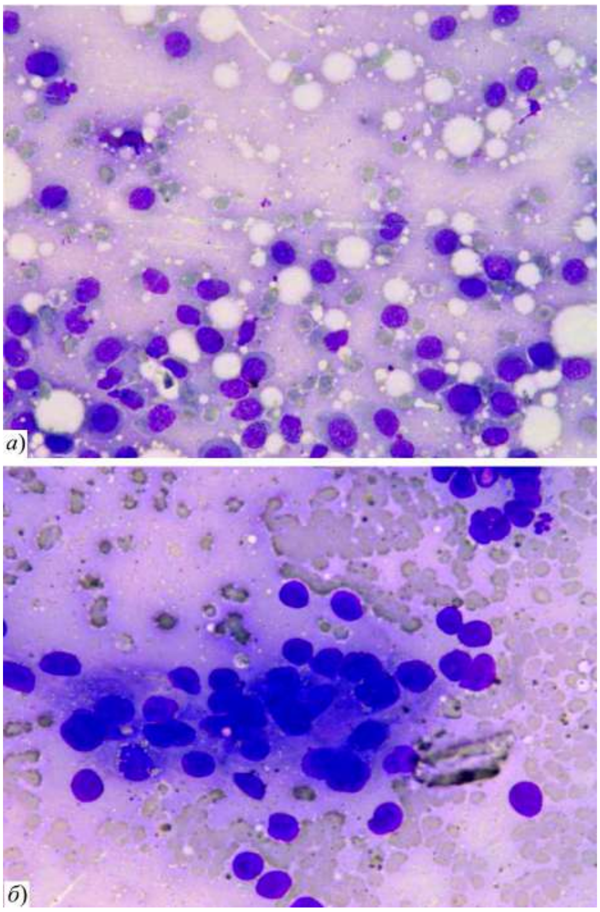

Рис. 1. Цитограма. Забарвлення за Романовським-Гімзою. Долькова непроліферативна мастопатія (*а*), фіброзна непроліферативна мастопатія (*б*). Збільшення ×200

Ознакою проліферації клітин можуть бути зміни морфологічних властивостей клітин або поява незвичних для незміненого органа багатоклітинних епітеліальних структур. До клітинних критеріїв проліферації належать зменшення розмірів клітин і їх ядер. У таких випадках переважають округлі або злегка овальні (ніби набухлі) ядра з чітко визначеним ніжнопетлистим інтенсивно зафарбованим хроматином. В окремих ядрах можуть визначатися поодинокі, порівняно невеликі, інтенсивно зафарбовані ядерця. Цитоплазма переважної більшості клітин зафарбована в інтенсивно голубі тони, іноді бузкові або сині. Клітинні комплекси, що характеризують проліферацію, мають вигляд округлих утворень, подібних до первинних залозистих структур (ацинусів) без сформованих у них просвітів – збільшення кількості клітин. Трапляються також сосочкоподібні комплекси зі щільним розташуванням клітин у багатошарові пласти. Якщо у патологічному вогнищі переважає фіброз, то в цитологічних препаратах виявляється мала кількість мілких, мономорфних клітин з однорідними круглими гіперхромними ядрами і ледве помітним вузьким обідком інтенсивно зафарбованої цитоплазми (рис. 1, *б*). У результаті аналізу морфологічних ознак сформульовано правило для діагностування кістозної непроліферативної мастопатії:

ЯКЩО клітини розташовані пластами.

І трапляються кубічні та призматичні елементи, сосочкові та округлі комплекси,

І у фоні багато фагоцитів та гістіоцитів,

І в клітинах округле ядро розташовується центрально,

І трапляються клітини із апокриновою секрецією, які мають дві зони (базальну та апікальну),

ТО це кістозна непроліферативна мастопатія (80 %).

Подамо це правило у вигляді логічного правила *IF-THEN*.

Позначимо антецеденти:

A_1 = {*клітини розташовані пластами*};

A_2 = {*трапляються кубічні та призматичні елементи, сосочкові та округлі комплекси*};

A_3 = {*у фоні багато фагоцитів та гістіоцитів*};

A_4 = {*в клітинах округле ядро розташовується центрально*};

A_5 = {*трапляються клітини із апокриновою секрецією, які мають дві зони* (*базальну та апікальну*)}.

Консеквент буде такий:

K = {*кістозна непроліферативна мастопатія* (*80 %*)}.

Тоді правило матиме такий вигляд:

IF (A_1 & A_2 & A_3 & A_4 & A_5) *THEN 0,8 K.*

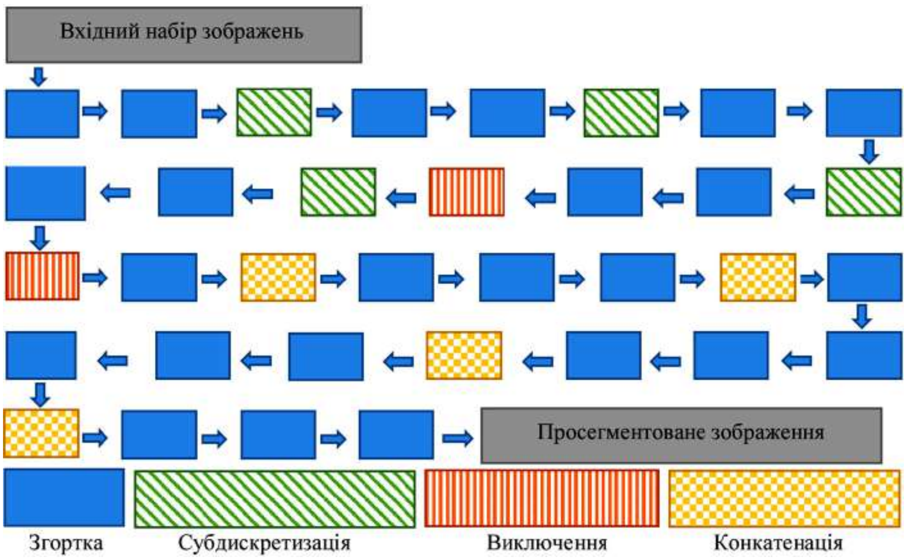

Рис. 3. Архітектура нейронної мережі типу U-Net

Метод опрацювання цитологічних зображень. Результатом мікроскопічного дослідження є цифрове зображення у форматах *.jpg, *.png, *.bmp. Для застосування алгоритмів лінійної регресії необхідно обчислити кількісні характеристики мікрооб’єктів. Кількісними характеристиками ядер клітин є площа, периметр, окружність. Метод опрацювання цитологічних зображень складається із таких етапів (рис. 2).

На першому етапі здійснюється завантаження зображення у пам’ять комп’ютера. На другому етапі відбувається попереднє оброблення зображень, що передбачає фільтрацію гауссовим та медіанним фільтрами для зниження рівня шумів. Додатково застосовують гістограмне вирівнювання.

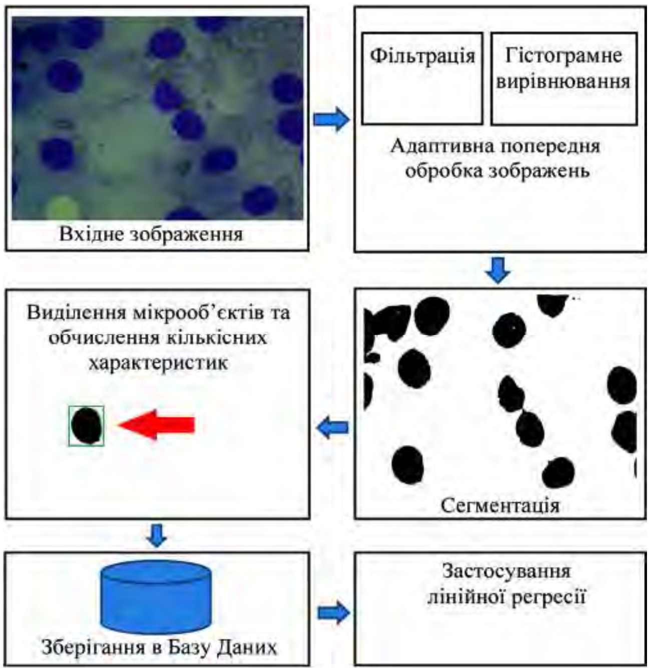

Рис. 2. Етапи опрацювання цитологічних зображень

На третьому етапі відбувається автоматична сегментація зображень на підставі згорткової нейронної мережі типу U-Net. На четвертому етапі здійснюється обчислення кількісних характеристик ядер клітин, виділених після етапу сегментації. На п’ятому етапі кількісні характеристик ядер клітин зберігаються у базі даних для подальшого оброблення.

На останньому етапі застосовуються алгоритми лінійної регресії для визначення кількісних характеристик ядер клітин. Для проведення автоматичної сегментації розроблено архітектуру згорткової нейронної мережі типу U-Net (рис. 3).

Згорткові нейронні мережі – вид нейронних мереж, які хоча б на одному зі своїх шарів для перетворення використовують операцію “згортки”. Згортковий шар є набором карт (карти ознак). У кожній карті є синаптичне ядро (скануюче ядро або фільтр). Розмір карт обчислюють за формулою:

$$(w,h) = (mW - kW + 1, mH - kH + 1),$$

де (w,h) – розмір карти згортки; mW – ширина попередньої карти; mH – висота попередньої карти; kW – ширина ядра; kH – висота ядра.

Якщо на вході є двовимірне зображення I та ядро K, то операція згортки має такий вигляд:

$$s(i,j) = \langle I * K \rangle (i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n).$$

Формально шар субдискретизації можна описати такою формулою:

$$x^l = f\left(a^l \cdot subsample\left(x^{l-1}\right) + b^l\right),$$

де x^l – вихід шару l; $f()$ – функція активації; a^l, b^l – коефіцієнти зсуву шару l; $subsample()$ – операція вибору локальних максимальних значень.

Виключення (Dropout) – це метод регуляризації нейронних мереж, призначений для запобігання перенавчанню мережі. Цей шар характеризує вилучення певного відсотка випадкових нейронів (що містяться як у прихованих, так і у видимих шарах) на різних ітераціях (епохах) під час навчання нейронної мережі. Dropout вилучає нейрони з ймовірністю p, як наслідок, залишає їх включеними з ймовірністю $q = 1 - p$.

Важливим аспектом U-Net є поєднання низхідного та висхідного шляхів за допомогою конкатенацій. Роблячи це, мережа може навчитися класифікувати та локалізувати, використовуючи наскрізний підхід до навчання. Для оцінки регресії використано похибку, яка дорівнює квадратному кореню із середньоквадратичної похибки (RMSE).

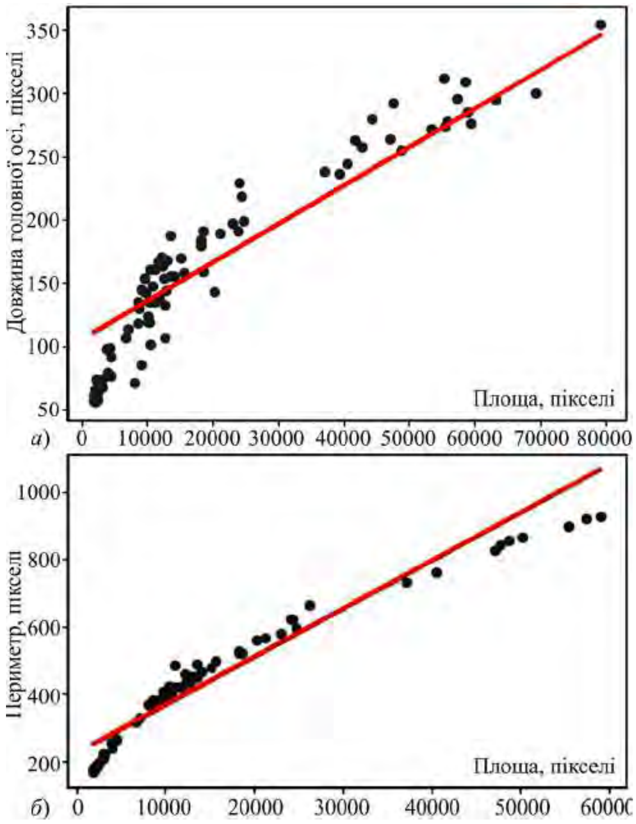

Рис. 4. Результат застосування лінійної регресії для параметрів: "площа, довжина головної осі" (*а*), "площа, периметр" (*б*)

Обговорення результатів дослідження. Рис. 4 та 5 ілюструють результати дослідження. На рис. 4, *а* наведено результат застосування лінійної регресії для параметрів "площа, довжина головної осі".

Значення RMSE для параметрів "площа, довжина головної осі" становить 26.15. На рис. 4, *б* наведено результат застосування лінійної регресії для параметрів "площа, периметр". Значення RMSE для параметрів "площа, периметр" становить 38.21. Результати порівняння параметрів кількісних характеристик мікрооб'єктів засобами лінійної регресії наведено на рис. 5.

Виконавши порівняння, можна зробити висновок про те, що використання сукупності параметрів "площа, довжина головної осі" дає меншу похибку RMSE порівняно із сукупністю параметрів "площа, периметр".

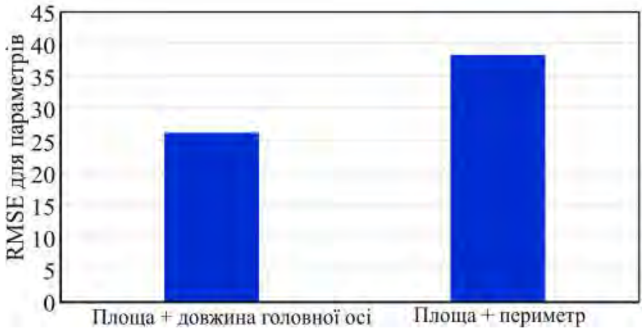

Рис. 5. Результати порівняння параметрів кількісних характеристик мікрооб'єктів на підставі лінійної регресії

У цій роботі, на відміну від проаналізованих наукових праць [1], [2], [7], [8], [10], [11], [12], [13], детально проаналізовано та наведено усі етапи опрацювання зображень: фільтрацію, попереднє оброблення, сегментацію на підставі згорткових нейронних мереж, виділення кількісних характеристик ядер клітин і застосування лінійної регресії.

Висновки

Розроблено метод опрацювання цитологічних зображень на підставі алгоритмів фільтрації та автоматичної сегментації зображень, що дало можливість визначати кількісні характеристики ядер клітин. На підставі згорткових мереж розроблено архітектуру мережі U-Net для здійснення сегментації цитологічних зображень.

На підставі визначених кількісних характеристик ядер клітин, завдяки використанню методу лінійної регресії, спрогнозовано кількісні показники мікрооб'єктів на цитологічному зображенні. На підставі проведених комп'ютерних експериментів виявлено, що сукупність параметрів "площа, довжина головної осі" дає в 1,4 разу меншу похибку RMSE порівняно із сукупністю "площа, периметр".

Подяка. Співавтор цієї роботи О. М. Березький почав займатися біомедичними зображеннями, починаючи з міжнародного проекту УНТЦ 1702 "Інформаційний відеоскоп надвисокої роздільної здатності для дослідження запрограмованої смерті (апоптозу) клітин пухлин людини", керівником якого був член-кореспондент НАНУ проф. В. В. Грицик. Нині в Західноукраїнському національному університеті працює наукова група, яка реалізує "Аналіз біомедичних зображень". Її очолює професор О. М. Березький.

Низький уклін шановному учителю!

References

[1] Abdulqader, Q. (2017). Applying the Binary Logistic Regression Analysis on The Medical Data. *Science Journal of University of Zakho*, *5*(4), 330–334. https://doi.org/10.25271/2017.5.4.388

[2] Altman, N. & Krzywinski, M. (2015). Simple linear regression. *Nat Methods*, *12*, 999–1000. https://doi.org/10.1038/nmeth.3627

[3] Berezsky, O. M. (Ed.) (2017). Methods, algorithms and software for processing biomedical images. Ternopil: Ekonomichna dumka, TNEU, 330. [In Ukrainian].

[4] Berezsky, O. M., Melnyk, G. M. & Berezka, K. M. (2013). Fuzzy knowledge base of the intelligent system for diagnosing breast cancer. *Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky, 6*, 284–291. [In Ukrainian].

[5] Berezsky, O. M., Melnyk, G. M., Batko, Y. M., & Datsko, T. V. (2013). Intelligent system for diagnosing various forms of breast cancer based on the analysis of histological and cytological images. *Scientific Bulletin of UNFU, 23*(13), 357–367. [In Ukrainian].

[6] Berezsky, O., Pitsun, O., Dubchak, L., Berezka, K., Dolynyuk, T., & Derish, B. (2020). Cytological Images Clustering of Breast Pathologies. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, Zbarazh, Ukraine, 62–65. https://doi.org/10.1109/CSIT49958.2020.9321867

[7] Deepa, S. N. & Aruna Devi, B. (2011). A survey on artificial intelligence approaches for medical image classification. *Indian Journal of Science and Technology, 4*(11), 1583–1595. https://doi.org/10.17485/ijst/2011/v4i11.35

[8] Dreiseitl, S. & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics, 35*(5–6), 352–359. https://doi.org/10.1016/s1532-0464(03)00034-0

[9] Fabijańska, A. & Sankowski, D. (2011). Noise adaptive switching median-based filter for impulse noise removal from extremely corrupted images. *IET Image Processing, 5*(5), 472–480. https://doi.org/10.1049/iet-ipr.2009.0178

[10] Omer, A. A., Hassan, O. I., Ahmed, A. I. & Abdelrahman, A. (2018). Denoising CT Images using Median based Filters: a Review. International Conference on Computer, Control, Electrical and Electronics Engineering (ICCCEEE). Khartoum, Sudan, 1–6. https:// doi.org/10.1109/ICCCEEE.2018.8515829

[11] Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., & Lin, C.-W. (2020). Deep learning on image denoising: An overview. *Neural Networks, 131*, 251–275. https:// doi.org/10.1016/j.neunet.2020.07.025

[12] Zhang, J., Xiea, Y., Wu, Q. & Xia Y. (2019). Medical image classification using synergic deep learning. *Medical Image Analysis, 54*, 10–19. https://doi.org/10.1016/j.media.2019.02.010

[13] Zhu, Y., & Huang, C. (2012). An Improved Median Filtering Algorithm for Image Noise Reduction. *Physics Procedia, 25*, 609–616. https://doi.org/10.1016/j.phpro.2012.03.133

O. M. Berezsky[1], O. Yo. Pitsun[1], G. M. Melnyk[1], T. V. Datsko[2]

[1] *West Ukrainian National University, Ternopil, Ukraine*
[2] *Ternopil National Medical University, Ternopil, Ukraine*

APPLICATION OF LINEAR REGRESSION METHOD FOR ANALYSIS OF CYTOLOGICAL IMAGES QUANTITATIVE CHARACTERISTICS

This article analyzes the pathological conditions of the breast based on the study of cytological images. Cytological images are a separate class of biomedical images and are used in the diagnosis of cancer. For diagnose precancerous and cancerous conditions and treatment tactics, diagnosticians use cytological, histological, and immunohistochemical images. For automating the process of diagnosis in oncology, automated microscopy systems are used. Automated microscopy systems use computer vision algorithms. Recently, machine learning algorithms have been used to classify images. Microscopic image processing is a complex and time-consuming process, as the images are characterized by high noise levels and the absence of clear contours of cell nuclei. To calculate the quantitative characteristics of cell nuclei cytological images, the method for calculating the quantitative characteristics of cell nuclei based on image filtering algorithms and their automatic segmentation has been developed. An U-Net convolutional neural network architecture has been developed for cell nucleus segmentation. In this work, the method of processing cytological images is developed. The method consists of six stages. The first step is to load the image into the computers memory. In the second stage, the images are preprocessed. The third stage is the automatic segmentation of images based on the convolutional neural network of the U-Net type. In the fourth stage, the quantitative characteristics of cell nuclei are calculated. In the fifth stage, the quantitative characteristics of the cell nuclei are stored in a database. In the sixth stage, linear regression algorithms are used to obtain quantitative characteristics of cell nuclei. Currently, linear regression is one of the common approaches of machine learning to data analysis. In this work, the comparative analysis of the quantitative characteristics application in cell nuclei is carried out based on linear regression. The scientific novelty of the work is development the method for calculating the quantitative characteristics of cell nuclei, which includes stages of image filtering and automatic segmentation based on the use of a neural network such as U-Net. The practical significance of the work is the software implementation of the preprocessing modules and linear regression. In particular, investigated that the set of parameters “area, length of the main axis” has 1.4 times less RMSE error compared to the set “area, perimeter”.

Key words: automatic segmentation; cytological images; U-Net neural network; preprocessing; linear regression.

Інформація про авторів:

Березький Олег Миколайович, д-р техн. наук, професор, завідувач кафедри комп’ютерної інженерії. **Email:** olber62@gmail.com; https://orcid.org/0000-0001-9931-4154; ResearcherID: I-7196-2017

Піцун Олег Йосипович, ст. викладач, кафедра комп’ютерної інженерії. **Email:** o.pitsun@wunu.edu.ua; https://orcid.org/0000-0003-0280-8786

Мельник Григорій Миколайович, канд. техн. наук, кафедра комп’ютерної інженерії. **Email:** mgm@wunu.edu.ua; https://orcid.org/0000-0003-0646-7448

Дацко Тамара Вікторівна, канд. мед. наук, доцент, кафедра патологічної анатомії з секційним курсом та судовою медициною. **Email:** datsko_t@tdmu.edu.ua; https://orcid.org/0000-0001-9283-2629

Цитування за ДСТУ: Березький О. М., Піцун О. Й., Мельник Г. М., Дацко Т. В. Застосування методу лінійної регресії для аналізу кількісних характеристик цитологічних зображень. Український журнал інформаційних технологій. 2021, т. 3, № 1. С. 73–77.

Citation APA: Berezsky, O. M., Pitsun, O. Yo., Melnyk, G. M., & Datsko, T. V. (2021). Application of linear regression method for analysis of cytological images quantitative characteristics. *Ukrainian Journal of Information Technology*, *3*(1), 73–77. https://doi.org/10.23939/ujit2021.03.073