

Классификация аномалий молочной железы с использованием глубокого обучения

П.С. Гомина*, В.И. Кобер**, В.Н. Карнаухов**, М.Г. Мозеров**, А.В. Кобер***

* Челябинский государственный университет, Челябинск, Россия

** Институт проблем передачи информации, Российская академия наук, Москва, Россия

*** Федеральный научный центр биологических систем и агротехнологий РАН, Оренбург, Россия

Поступила в редколлегию 14.06.2022

Аннотация—Раннее обнаружение аномалий молочной железы с помощью маммографического скрининга и правильное лечение снижает смертность и увеличивает продолжительность жизни женщин. В настоящее время активно разрабатываются методы и алгоритмы для систем компьютерной диагностики на основе глубоких нейронных сетей. Такие системы объединяют в себе как выбор, расчет признаков, так и классификацию, тем самым напрямую создавая функцию принятия решений. В данной работе предлагается метод классификации патологий молочной железы в соответствии с международной системой интерпретации и протоколирования визуализации молочной железы (BI-RADS), основанный на глубоком обучении. Экспериментальные результаты представлены с использованием двух открытых баз данных цифровой маммографии и оценены с использованием общепринятых критериев качества.

КЛЮЧЕВЫЕ СЛОВА: сегментация, цифровая маммография, глубокая сверточная нейронная сеть U-net, аугментация данных

DOI: 10.53921/18195822_2022_22_2_111

ВВЕДЕНИЕ

По данным Всемирной Организации Здравоохранения, ежегодно рак молочной железы убивает более полумиллиона женщин во всем мире [1]. Однако раннее выявление рака молочной железы (РМЖ) и правильное лечение позволяют существенно увеличить продолжительность жизни. Несмотря на то, что существуют различные способы выявления рака молочной железы, такие как магнитно-резонансная томография, томосинтез, ультразвуковое исследование, рентгеновская маммография продолжает оставаться наиболее широко используемым способом раннего выявления патологий молочной железы. Цифровая маммография использует специальный электронный детектор рентгеновского излучения, который преобразует сигнал в цифровое изображение для компьютерного скрининга (визуальное обследование для выявления заболеваний). Для обнаружения патологий молочной железы используются две ее проекции: косая медиалатеральные (MLO) и левая/правая краниокаудальные (CC) проекции [2]. Стандартное описание маммограмм [3] включает в себя следующую информацию: показание к проведению обследования (скрининг; диагностическое исследование; контрольное наблюдение); описание структуры молочных желёз по рекомендации Американской Коллегии Радиологов (ACR); описание патологических признаков с использованием стандартной терминологии (объемное образование, асимметрия, нарушение архитектоники, кальцификация, сопутствующие признаки); сравнение с предыдущими исследованиями (при наличии); вывод/заключение с выставлением итоговой категории по шкале BI-RADS [4] — система интерпретации и протоколирования визуализации молочной железы. Основной целью создания системы BI-RADS была разработка унифицированного языка/терминов интерпретации результатов осмотра молочных желёз,

классификация выявляемых радиологических признаков и адекватных рекомендаций к дальнейшей лечению. По данным ACR объемное образование определяется как область поражения, видимая, по крайней мере, в двух разных проекциях. Объемные образования имеют разную плотность, края и форму [5]. Образования округлой и овальной формы с гладкими краями обычно указывают на доброкачественные изменения, а вот резкие и неровные границы указывают на злокачественное образование. Как правило, рентгенологические изображения нелегко интерпретировать. Так визуальная диагностика РМЖ по рентгеновским маммограммам имеет ложно-отрицательный (пропущенный рак) уровень более 16%, что связано с плотной тканью молочной железы, скрывающей рак, со зрительными ограничениями человека и с наличием шума на изображениях [6]. Для облегчения задачи врачей по обнаружению патологий молочной железы на ранних стадиях, в настоящее время активно разрабатываются методы и алгоритмы для систем компьютерной диагностики (СКД). Традиционный подход к созданию таких систем состоит из выбора и расчета признаков интересующих объектов из исходных данных с последующим подбором модели для их классификации. По сути, типичная СКД состоит из следующих последовательных этапов [7]:

(а) этап предварительной обработки, который используется для улучшения качества изображения, удаления шума и устранения меток на изображении маммограммы;

(б) этап сегментации областей интереса, который используется для извлечения областей, содержащих патологии;

(в) извлечение различных текстурных, статистических и геометрических признаков;

(г) этап классификации, где ранее обученный классификатор используется для выявления и классификации патологий в тканях молочной железы.

В работе [8] предложена система для диагностики рака молочной железы, основанная на вейвлет-преобразовании для получения признаков в области интереса. В целях уменьшения размера описания признаков используется гибридный подход к выбору признаков, в котором генетический алгоритм и машина опорных векторов объединены, чтобы обнаружить и классифицировать патологии. Отметим, что в современных системах компьютерной диагностики используются в основном линейные методы обработки [9, 10]. Так методы главных и независимых компонент являются лишь способами понижения размерности пространства признаков, вейвлет-преобразования — интегральные преобразования, смесь гауссовских распределений состоит из ограниченного количества гауссовских функций.

В последние годы были достигнуты значительные успехи в сложных задачах распознавания изображений благодаря применению глубоких нейронных сетей с тремя и более слоями, которые объединяют в себе как выбор и расчет признаков, так и классификацию, тем самым напрямую создавая функцию принятия решений. Очевидно, что чем больше слоев в таких сетях, тем более сложные нелинейные функции они способны создавать. Более того, глубокое обучение имеет преимущество перед традиционными методами машинного обучения в силу своей природы обучения: чем больше данных в обучающей выборке, тем точнее классификация [11]. Недавно были предложены эффективные глубокие сверточные нейронные сети (ГСНС) для решения задач классификации: VGGNet [12]. В 2014 году была предложена 19-слойная ГСНС. Сеть использовала очень маленькие сверточные фильтры 3x3 и показала значительное улучшение на базе данных ILSVRC. Было показано, что ГСНС должны иметь более глубокую сеть уровней, чтобы иерархические представления признаков работали хорошо. GoogLeNet [13]. В 2014 году также была предложена ГСНС, которая использовала параллельные структуры в своей архитектуре (9 модулей Inception с более чем 100 уровнями в общей сложности). ResNet [14]. В 2015 году была предложена ГСНС с 152-уровневую архитектурой. ResNet достигла минимальной ошибки (3:57%) в задаче классификации на базе данных ILSVRC. Структура остаточного обучения сети в 8 раз глубже, чем VGGNet, но все

же имеет меньшую вычислительную сложность. Все три ГСНС были разработаны для задачи классификации 1000 классов. ГСНС автоматически извлекают признаки из большого набора данных в процессе обучения для решения задач классификации. Недавние достижения в области глубоких нейронных сетей сделали возможным автоматическое обучение признаков на основе большого количества обучающих данных, обеспечивая комплексное решение от извлечения признаков до построения классификатора [15–19]. В данной работе использовалась модель сети ResNet-50 [20] для обнаружения и классификации патологий на маммограммах. Большинство баз данных маммограмм не являются общедоступными, что создает проблему для сравнения качества работы СДК. В данной работе были использованы две открытые базы данных - MIAS [21] и CSAW-M [22]. MIAS содержит 322 аннотированных маммограмм. База данных также включает в себя отметки радиолога о местонахождении всех патологий на изображениях. CSAW-M — крупнейшая аннотированная база данных, содержащая более десяти тысяч маммограмм. В этой работе предлагается метод классификации по шкале BI-RADS с использованием двух параллельно работающих ГСНС. Экспериментальные результаты представлены с использованием известных баз данных по цифровой маммографии. Статья организована следующим образом: в разделе 1 описан предлагаемый метод классификации, в разделе 2 представлены и обсуждены полученные результаты экспериментов с помощью компьютерного моделирования и раздел 3 суммирует наши выводы.

1. BI-RADS КЛАССИФИКАЦИЯ АНОМАЛИЙ МОЛОЧНОЙ ЖЕЛЕЗЫ

На первом этапе устраняются шум и метки, улучшается качество изображений, которые использовались в обучении и тестировании [23–26]. Все изображения подверглись выравниванию гистограммы для повышения контрастности. Затем изображения были нормализованы путем вычитания среднего значения из каждого элемента и деления результата на среднеквадратичное отклонение. Все изображения левой груди были перевернуты таким образом, чтобы грудь всегда находилась на одной стороне изображения. Исходные изображения в базах данных различаются по размеру. Так изображения базы данных MIAS имеют размер 1024x1024 пикселей, в то время как изображения базы данных CSAW-M имеют размер 632x512 пикселей. Все изображения базы данных MIAS были приведены к размеру базы данных CSAW-M путем обрезания фона и небольшого масштабирования. Отметим, что многие исследования по классификации изображений и распознаванию объектов естественных изображений уменьшают исходные изображения с высоким разрешением, чтобы улучшить использование памяти и уменьшить вычислительную сложность своих моделей. Например, одна из самых эффективных моделей для задачи классификации ImageNet обучена по изображениям, уменьшенным до 224x224 пикселей. Понижающая выборка хорошо работает для естественных сцен, где грубые признаки, такие как формы, цвета и другие глобальные дескрипторы, используются для выполнения задач классификации. Однако сильное уменьшение размера маммографических изображений может привести к потере тонких структур и мелких деталей, которые могут свидетельствовать о раннем раке. В соответствии с международной системой интерпретации и протоколирования визуализации молочной железы [4], выделяют следующие категории по шкале BI-RADS:

Категория 0. Невозможно прийти к однозначному выводу по результатам визуализации (маммография плохого качества, неправильная укладка, недостаточно проекций).

Категория 1. Маммограммы без патологических образований, нарушений архитектоники или подозрительных кальцинатов, допустимы неизменные аксиллярные лимфатические узлы (возрастная норма). Вероятность злокачественности 0%.

Категория 2. Доброкачественные изменения, такие как любые диффузные формы фиброзно-кистозная мастопатия; простая киста без признаков воспаления; инволютивная фиброаденома;

доброкачественные обызвествления, сосудистые кальцинаты; диффузные кальцинаты; интрамаммарный лимфатический узел; дуктэктазия; липома/фибролипома; гамартома; галактоцеле; оперированный рак молочной железы без признаков рецидива; отек после лучевой терапии. Вероятность злокачественности 0%.

Категория 3. Доброкачественные изменения, такие как впервые выявленные киста, фиброаденома; непальпируемое образование с четкими ровными контурами, единичная группа точечных микрокальцинатов; мастит; втяжение соска. Вероятность злокачественности 0–2%.

Категория 4. Изменения, подозрительные на злокачественный процесс, необходима морфологическая верификация. Используется в случае обнаружения изменений, не имеющих всех признаков злокачественности, но крайне подозрительных в отношении малигнизации, а также при односторонней подозрительной лимфоаденопатии без четко определяемой патологии в молочной железе. Эта категория имеет широкий диапазон вероятности злокачественности (2–95%).

Категория 5. Достоверно злокачественные изменения. Выставляется в случае обнаружения изменений, рентгенологически имеющее все признаки злокачественности: объемное образование высокой плотности с неровным лучистым контуром; сегментарное или линейное распределение мелких линейных и плеоморфных кальцинатов; объемное образование с неровным лучистым контуром в сочетании с плеоморфными кальцинатами. Вероятность злокачественности (>95%).

Категория 6. Гистологически подтвержденный РМЖ. Злокачественность доказана.

В данной работе проводилась классификация маммограмм на первые три категории BI-RADS: категория 0, категория 1 и категория 2. Модель системы классификации состоит из двух ГСНС, которые обрабатывают одновременно проекции MLO и CC. В каждой ветви модели сеть ResNet-50 извлекает признаки, которые после усреднения (для уменьшения размерности векторов признаков) объединяются в единый вектор [27]. Этот вектор подается на полносвязный слой, за которым следует классификация с помощью обобщенной логической функции softmax, вычисляющей выходное распределение $p(y/x)$. Схема модели предлагаемой классификации с использованием сверточных нейронных сетей показана на Рис. 1.

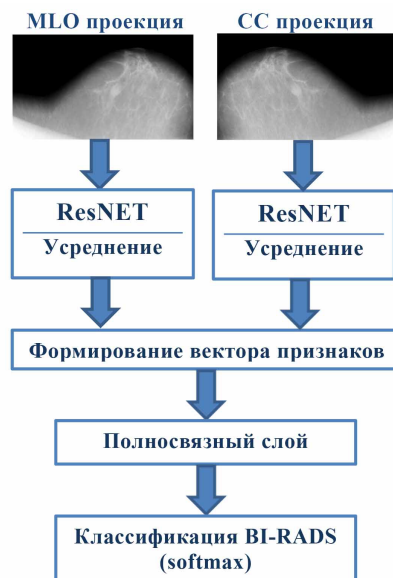


Рис. 1. Схема BI-RADS классификации с помощью глубокой сверточной нейронной сети.

Для обучения модели использовался размер пакета 64 и изображения 632x512 пикселей с помощью оптимизатора Adam [28]. Инициация производилась предварительно обученными весами ImageNet [29]. Скорость обучения выбрана 10^{-6} . Для увеличения данных при обучении применялось случайное горизонтальное и вертикальное отражение, а также случайные повороты на 10–15 градусов. Мы использовали 5-кратную перекрестную проверку, чтобы определить итерацию останова, которая использовалась в последнем обучающем прогоне.

2. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Для оценки эффективности предложенного подхода были использованы следующие критерии: точность, чувствительность и специфичность. Эти метрики определяются следующим образом [8]:

$$accuracy = \frac{TP + TN}{TN + FP + FN + TP}, \quad (1)$$

$$specificity = \frac{TN}{TN + FP}, \quad (2)$$

$$sensitivity = \frac{TP}{FN + TP}, \quad (3)$$

где TN — истинно-отрицательный, FP — ложно-положительный, FN — ложно-отрицательный и TP — истинно-положительный результаты экспериментов. Чувствительность и специфичность — это статистические показатели эффективности теста бинарной классификации. Оба показателя являются относительными, выражаются в долях единицы или процентах. Чувствительность (истинно положительная пропорция) отражает долю положительных результатов, которые правильно идентифицированы как таковые (то есть, чувствительность диагностического теста показывает вероятность того, что больной пациент будет классифицирован именно как больной). Специфичность (истинно отрицательная пропорция) отражает долю отрицательных результатов, которые правильно идентифицированы как таковые (то есть, вероятность того, что не больные пациенты будут классифицированы именно как не больные). Качество бинарной классификации может быть оценено с помощью графика ошибок ROC-кривая (Receiver Operating Characteristic), который отображает соотношение между чувствительностью и специфичностью алгоритма классификации при варьировании порога решающего правила. На основе этой кривой строится один из самых популярных функционалов качества в задачах бинарной классификации — площадь под ROC-кривой, обозначаемый AUC (Area Under the Curve). Чем выше показатель AUC, тем качественнее классификатор, Базы данных CSAW-M и MIAS использовались для обучения предложенной модели, а для оценки классификации использовались только изображения базы данных CSAW-M. Мы построили три разных графика ROC для трех категорий BI-RADS (0, 1 и 2) с использованием методологии бинарной классификации “один против всех” и вычислили соответствующие значения AUC. Результаты экспериментов приведены в Таблице 1.

Таблица 1. Качество классификации тестовых маммограмм по шкале BI-RADS.

	Точность %	Специфичность %	Чувствительность %	AUC
Категория 0	83.5	89.2	90.2	0.88
Категория 1	81.3	88.6	89.4	0.86
Категория 2	81.2	88.4	89.2	0.84

Отметим, что предложенная схема обработки достигает наилучшей классификации для BI-RADS категории 0. Однако для дальнейшего улучшения качества классификации необходимо существенно увеличить обучающую выборку, которая должна состоять не менее чем из нескольких десятков тысяч маммограмм [30].

3. ЗАКЛЮЧЕНИЕ

В данной работе предложен метод классификации аномалий молочной железы по шкале BI-RADS с использованием двух входных проекций (MLO и CC) с высоким разрешением и модели глубокого обучения, построенной на основе двух параллельных глубоких сверточных нейронных сетей. Полученные экспериментальные результаты показывают, что предложенная модель обладает высоким качеством классификации всех трех тестируемых категорий с точки зрения показателя AUC. Однако эффективность предложенной модели можно существенно улучшить путем увеличения обучающей выборки, которая должна состоять не менее чем из нескольких десятков тысяч маммограмм.

Благодарность. Исследование поддержано грантом Российского научного фонда (проект № 22-19-20071).

СПИСОК ЛИТЕРАТУРЫ

1. Bray F., Ferlay J., Soerjomataram I., Siegel R., Torre L., Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A cancer journal for clinicians*, 2018, vol. 68(6), pp. 394–424, 2018.
2. Методические рекомендации по использованию системы BI-RADS при маммографическом обследовании. Под ред. А. Ю. Васильева. Москва. 23с., 2017.
3. D’Orsi C.J., Sickles E.A., Mendelson E.B., Morris E.A. et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Fifth Edition. Reston VA, American College of Radiology, 2013, pp. 1-79.
4. American College of Radiology. ACR Breast Imaging Reporting and Data System (BI-RADS) Website, www.acr.org.
5. Bozek J., Mustra M., Delac K., Grgic M. A survey of image processing algorithms in digital mammography. *Recent advances in multimedia signal processing and communications*, 2009, pp. 631–657.
6. Trister A.D., Buist D., Lee C. Will machine learning tip the balance in breast cancer screening? *JAMA oncology*, 2017, vol. 3(11), pp. 1463–1464.
7. Luqman Mahmood Mina and Nor Ashidi Mat Isa. A review of computer-aided detection and diagnosis of breast cancer in digital mammography. *Journal of Medical Sciences*, 2015, vol. 15(3), pp. 110.
8. Mohamed S. Salama, Ahmed S. Eltrass, Hassan M. Elkamchouchi. An improved approach for computer-aided diagnosis of breast cancer in digital mammography. *IEEE International Symposium on Medical Measurements and Applications*, 2018, pp. 1–5.
9. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*, 2017, vol. 31 (4), pp. 198-211.
10. Halalli D., Makandar A. Computer Aided Diagnosis - Medical Image Analysis Techniques, Breast Imaging. C. M. Kuzmiak (Ed.), IntechOpen, 2018.
11. Goodfellow Y., Bengio A. Courville. *Deep Learning*. MIT Press, 2016.
12. Simonyan K., Zisserman F., Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.
13. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. *IEEE Computer Vision and Pattern Recognition*, 2015. <http://arxiv.org/abs/1409.4842>.

14. He K., Zhang X., Ren S., Sun, J. Deep residual learning for image recognition. Tech. report, 2015. <http://arxiv.org/abs/1512.03385>.
15. Zou L., Yu S., Meng T., Zhang Z., Liang X., Xie Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Computational and Mathematical Methods in Medicine*, 2019, vol. 219, pp. 6509357.
16. Wang H., Feng J., Zhang Z. et al. Breast mass classification via deeply integrating the contextual information from multi-view data. *Pattern Recognition*, 2018, vol. 80, pp. 42–52.
17. Dhungel N., Carneiro G., Bradley A.P. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical Image Analysis*, 2017, vol. 37, pp. 114–128.
18. Saeed Seyyedi, Margaret J. Wong, Debra M. Ikeda, Curtis P. Langlotz. SCREENet: A Multi-view Deep Convolutional Neural Network for Classification of High-resolution Synthetic Mammographic Screening Scans. arXiv:2009.08563v3, 2020.
19. Ruchai A., Kober V., Dorofeev K., Karnaukhov V., Mozerov M. Classification of breast abnormalities using a deep convolutional neural network and transfer learning. *Journal of Communications Technology and Electronics*, 2021, vol. 66(6), pp. 778-783.
20. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. arXiv:1512.03385v12015.
21. Suckling J., et al. The mammographic image analysis society digital mammogram database. Mammographic Image Analysis Society (MIAS) database v1.21 [Dataset], 2015. <https://www.repository.cam.ac.uk/handle/1810/250394>.
22. Sorkhei M., Liu Y., Azizpour H. et al. CSAW-M: An Ordinal Classification Dataset for Benchmarking Mammographic Masking of Cancer. arXiv:2112.01330v1, 2021.
23. Diaz-Escobar J., Kober V., Karnaukhov V., Mozerov M. Recognition of breast abnormalities using phase features. *Journal of Communications Technology and Electronics*, 2020, vol. 65(12), pp. 1476-1483.
24. Kober V. Robust and efficient algorithm of image enhancement. *IEEE Transactions on Consumer Electronics*, 2006, vol. 52(2), pp. 655–659.
25. Kober V., Mozerov M., Alvarez-Borrego J. Nonlinear filters with spatially-connected neighborhoods. *Optical Engineering*, 2001, vol. 40(6), pp. 971-983.
26. Moreno I., Kober V., Lashin V., Campos J., Yaroslavsky L., Yzuel M. Color pattern recognition with circular component whitening. *Optics Letters*, 1996, vol. 21(7), pp.498-500.
27. Abdar M, Makarenkov V. Cwv-bann-svm ensemble learning classifier for an accurate diagnosis of breast cancer. *Measurement*, 2019, vol. 146, pp. 557–570.
28. Kingma D.P., Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.
29. Deng J., Dong W., Socher R., et al.: Imagenet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition*. 2009.
30. Dutta S., Gros E. Evaluation of the impact of deep learning architectural components selection and dataset size on a medical imaging task. *Medical Imaging: Imaging Informatics for Healthcare, Research, and Applications*, 2018.

Classification of breast abnormalities using deep learning

Gomina P., Kober V., Karnaukhov V., Mozerov M., Kober A.

Early detection of breast abnormalities through mammography screening and proper treatment reduces mortality and increases women's life expectancy. Currently, methods and algorithms for computer diagnostic systems based on deep neural networks are being actively developed. Such systems combine selection, feature calculation, and classification, thereby directly creating a decision-making function. In this paper, a method for classifying breast pathologies according to the International Breast Imaging Interpretation and Recording System (BI-RADS) based on deep learning is proposed. Experimental results are presented using two open databases of digital mammography and evaluated using various performance criteria.

KEYWORDS: breast abnormality, Computer Aided Detection (CAD), deep learning, classification, convolutional neural network.