Подготовка изображений маммограммы пациента для сегментации

Э. Ч. Угвуеке [1], Д. Палогианнидис [2]
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
[1] ugwuekeestherc@gmail.com, [2]dimitris.palogiannidis@gmail.com

***Аннотация*. Раннее выявление рака молочной железы улучшает показатели выживаемости пациентов и снижает показатели смертности и заболеваемости во всем мире. Для выявления рака молочной железы было использовано несколько методов визуализации, однако изображения маммограммы доказали свою эффективность для выявления его самых ранних стадий. Сегментация изображений имеет решающее значение для анализа изображений, поскольку она помогает классифицировать ткань молочной железы на различные сегменты, делая каждый сегмент более доступным для дальнейшего анализа. Цель этой работы – подготовить изображения маммограмм из китайской базы данных маммографии для сегментации изображений.**

Ключевые слова: обнаружение микрокальцификации, маммография, китайская база данных маммографии, подготовка изображений, сегментация

I. Вступление

Микрокальцификации (МК) – это крошечные отложения кальция, которые на маммограммах проявляются в виде более ярких пятен, чем нормальные ткани молочной железы [1, 2]. МК, диаметр которых варьируется от 0,1 до 1 мм и имеют узловатую форму, являются одним из основных ранних признаков или симптомов рака молочной железы, наблюдаемых на маммограммах [3, 4]. В зависимости от их морфологии (размера, плотности, формы и распределения) МК можно разделить на две категории: доброкачественные и злокачественные. Доброкачественные МК не являются злокачественными, имеют почти однородную круглую форму, диффузно рассеяны в тканях молочной железы, в то время как злокачественные МК имеют различную плотность, непредсказуемые формы и распределение и с большей вероятностью перерастают в рак молочной железы [2, 5].

Рак молочной железы и микрокальцификации тесно связаны. Скопления МС, являясь одной из наиболее часто встречающихся маммографических аномалий, приводят к повышенному риску развития рака молочной железы [1, 4]. Примерно 31–50 % непальпируемых поражений, обнаруженных во время скрининговой маммографии, являются МС, которые преобладают в большинстве случаев протоковой карциномы in Situ (DCIS) [6, 7]; следовательно, крайне важно обнаружить их как можно скорее, чтобы увеличить вероятность выживания при раке молочной железы [8].

Маммограмма – это двумерное изображение, которое помогает определить аномалии рака молочной железы (асимметричные кальцификации, шишки и деформированные участки молочной железы), которые морфологически сомнительны. Маммографическая визуализация использует низкоэнергетические рентгеновские лучи с энергией 20–30 кэВ и в настоящее время является общепринятым и стандартным подходом к скринингу и диагностике [9]. Это один из наиболее устоявшихся и широко используемых методов скрининга молочной железы, при котором для получения изображений молочной железы используется ионизирующее излучение в форме рентгеновского снимка [10]. С дополнительным преимуществом одновременного скрининга на рак молочной железы, не связанного с имеющимися симптомами, диагностическая маммография рекомендуется женщинам старше 40 лет с признаками и симптомами рака молочной железы [11].

Рентгенологи часто с трудом обнаруживают кальцификации, которые являются необнаружимыми аномалиями на маммограммах, которые могут быть злокачественными. Обнаружение кальцинатов на маммограммах – это тонкая и отнимающая много времени операция, которая напрягает зрение, снижает точность обнаружения и в конечном итоге приводит к ошибкам рентгенологов. Таким образом, эти крошечные пятна могут быть распознаны рентгенологами с помощью автоматизированного обнаружения (CADe). В настоящее время глубокое обучение (DL) используется в маммографии и других видах медицинских визуализационных исследований. До разработки DL для обнаружения поражений на маммограммах часто использовались традиционные методы машинного обучения (ML), но эти методы сложны и отнимают много времени, поскольку требуют разработки характеристик [12].

Обнаружение и классификация МС на изображениях маммографии с использованием обычных методов может серьезно повлиять на прогноз рака молочной железы и увеличить количество ложноположительных случаев. В этом исследовании объясняется алгоритм предварительной обработки изображений маммограммы, используемый для сегментации изображений и дальнейшей классификации с использованием глубокого обучения. В качестве входных данных для предварительной обработки использовались маммографические изображения, полученные из китайской базы данных маммографии.

II. Обзор литературы

Исследователи использовали несколько методов для обнаружения микрокальцинатов и их классификации. При анализе медицинской визуализации сегментация изображения является наиболее часто используемым методом, а также одной из наиболее важных задач в обработке изображений, которая включает в себя разделение изображения на несколько областей с аналогично связанными характеристиками [13].

Чаудхури С. и др. [14] предварительно обработали изображения с использованием алгоритма адаптивного выравнивания гистограммы с ограниченным контрастом (CLAHE) для удаления шума и повышения качества изображения перед сегментацией с помощью алгоритма K-средних. Чтобы определить их классификацию, они использовали такие методы, как машина нечетких опорных векторов (FSVM), байесовский классификатор и случайный лес. Результаты показали, что FSVM превзошел другие классификаторы по пяти параметрам производительности: точности, чувствительности, специфичности, прецизионности и отзывчивости на различных алгоритмах машинного обучения. Вместо того, чтобы сначала идентифицировать MCs отдельно, Ван Дж. и соавт. [15] разработали новую стратегию для выявления присутствия кластеризованных МС на маммограммах. Классификатор сверточной нейронной сети (CNN) получил кластеризованный МС непосредственно для классификации. На производительность классификации может повлиять снижение коэффициента истинного положительного результата из-за положения объекта, межкластерного разрыва и, возможно, шума между отдельными МС.

Существует необходимость в улучшении обнаружения и классификации МС, следовательно, компьютерное обнаружение имеет первостепенное значение для достижения этой цели и повышения уровня точности. Исследователи предложили несколько методов компьютерной диагностики для прогнозирования рака молочной железы на ранних стадиях по маммографическим изображениям с определенной точностью в процессе выявления. Важной целью этой статьи является обобщение предыдущих исследований глубоких сверточных нейронных сетей и кластеров МС для выявления рака молочной железы.

Сегментационный CNN был использован в [3] для определения местоположения МС на полных изображениях маммограммы без использования фрагментов изображения. Изображения были предварительно обработаны путем изменения размера и создания маски на изображениях с последующим увеличением изображения, чтобы уменьшить чрезмерную подгонку. После этого, используя сегментацию изображения и знания рентгенолога, была идентифицирована зона микрокальцификации. Были взяты пиксели микрокальцификации с положительной метки с помощью обучения сети сегментации. Используя архитектуру U-Net и сравнивая маску изображения с прогнозируемыми результатами их работы, результаты оценки автоматизированного обнаружения (CADe) с использованием метода среднеквадратичной логарифмической ошибки (MSLE) показали точность 95 % с погрешностью 5 %, в то время как метод количественной оценки достиг точности 90,3 %.

Кроме того, здесь упоминаются некоторые исследования, в которых использовались различные подходы к прогнозированию рака молочной железы и выявлению МС. Чтобы повысить точность классификации молочных желез, был создан CADe, использующий глубокую нейронную сеть для выявления подозрительных и нормальных паттернов на маммограммах [16]. Беккер и соавт. [17] протестировали модель глубокой нейронной сети (DNN) для цифровой маммографической диагностики рака молочной железы и достигли 81 % точности на ее максимальном уровне.

В исследовании, описанном в [18] с использованием DL, MobileNetV2 использовался для определения отличительных характеристик доброкачественного и злокачественного МС. Используя набор данных системы отчетов и данных о визуализации молочной железы (BI-RADS) категории 4, они объединили модели DL, одна из которых использовала исключительно маммографические изображения (CNN), а другая – оба маммографические изображения и клинические факторы (MobileNetV2). Результаты предложенной комбинированной модели показали благоприятные значения площади под кривой (AUC), чувствительности и специфичности в тестовом наборе данных 0,91, 85,3 % и 91,9% соответственно, что указывает на то, что модель может быть использована для стратификации риска бирадиации 4 МКЛ. Ван Дж. и соавт. [4] использовали DL для выявления рака молочной железы на основе МС.

Кануэй и соавт. [12] предложили новый метод обнаружения МС на изображениях маммограммы с использованием подхода Patch Learning Approach (PLA). Они использовали набор данных от 204 пациенток, прошедших скрининговую маммографию, которые предоставили в общей сложности 815 собственных маммограмм. Эксперты отметили наличие кальцинатов на маммограммах, в то время как определенное количество нормальных участков и участков, содержащих кальцификацию, было извлечено из каждой маммограммы после того, как она была разделена на участки фиксированного размера. Классификация участков обучающих данных на группы «нормальные» и «кальцифицированные» была достигнута с использованием ResNet-50 CNN. Затем производительность обученного CNN была оценена с использованием новых тестовых данных. Предложенная бинарная классификация пластырей с использованием предложенного PLA продемонстрировала точность классификации 96,7 %, чувствительность 96,7 % и специфичность 96,7 %.

Преобразование Хафа было использовано в [19], чтобы предложить полностью автоматизированный компьютерный метод обнаружения кластеров МС на цифровых маммограммах, а для категоризации использовался алгоритм кластеризации. В общей сложности 364 маммограммы от 182 пациенток были использованы для оценки эффективности обнаружения предложенного подхода, что дало истинно положительный коэффициент 91,78 %.

III. Материалы и методы

A. *Набор данных*

Китайская база данных маммографии была собрана из Архива визуализации рака (TC) [20], который состоит из 5202 маммографических изображений от 1775 пациентов. Набор данных включает левую и правую молочные железы с видами MLO (медиолатерально-косой) и CC (краниокаудальный), в общей сложности 1872 идентификатора пациентов, помеченных как доброкачественные или злокачественные с молекулярными подтипами для 749 пациентов. В табл. 1 приведено подробное описание количества злокачественных и доброкачественных патологий на основе клинических данных.

ТАБЛИЦА I. КОЛИЧЕСТВО ИЗОБРАЖЕНИЙ МАММОГРАММЫ С ДОБРОКАЧЕСТВЕННЫМИ И ЗЛОКАЧЕСТВЕННЫМИ ПАТОЛОГИЯМИ

		Ненормальность			
		Кальцификация	**Масса**	**Оба**	**Весь**
Классификация	**Доброкачественный**	64	412	80	556
	Злокачественный	198	737	381	1316
					1872
Молекулярные подтипы	**Люминал A**	18	96	38	152
	Люминал B	42	214	120	376
	HER2-обогащенный	31	42	62	135
	Тройной минус	7	65	14	86
					749

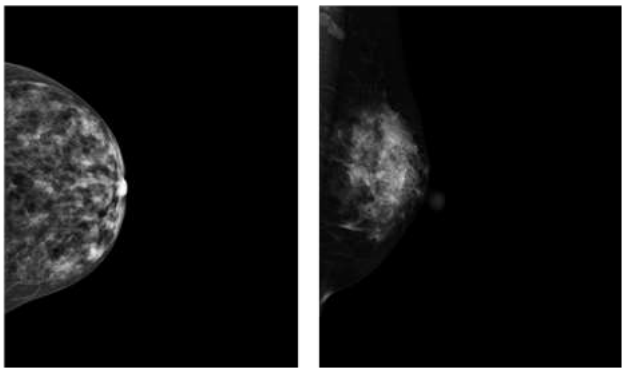
B. *Предварительная обработка изображений*

Изображения из базы данных были отсортированы путем извлечения изображений с аномалиями, помеченными как «Кальцификация» и «Оба» (кальцификация и масса), а затем снова отсортированы как «доброкачественная кальцификация» и «злокачественная кальцификация». Новый набор данных состоит из 144 доброкачественных и 579 злокачественных кальцинатов с разрешением 1914 x 2294, как показано на рис. 1.

Исходная маммограмма была получена и сохранена в формате Digital Imaging and Communications in Medicine (DICOM); поэтому для преобразования изображений в формат Portable Network Graphics (PNG) была использована автоматическая предварительная обработка, чтобы сохранить значения пикселей без ущерба для разрешения. Изображения маммограммы в формате PNG имели разрешение 1914 x 2294 пикселей, что позволило сохранить разрешение исходного изображения.

Этапы предварительной обработки изображения:

- Получите изображения DICOM.
- Прочитайте клинические данные и отсортируйте те, у которых есть отклонения.
- Отсортируйте изображения DICOM как «Доброкачественные» или «Злокачественные».
- Прочитайте изображения DICOM.
- Замените пиксельный массив DICOM на формат PNG.
- Верните изображение в формате PNG.
- Сохраните изображения в выходной каталог.



(а) Доброкачественный (б) Злокачественный

Рис. 1. Пример изображений маммограммы молочной железы из китайского набора данных маммограмм. (а) Краниокаудальный (CC) вид изображения доброкачественной кальцификации. (б) Медиолатеральный косой (MLO) вид изображения злокачественной кальцификации. Эти изображения были помечены клиническими данными в наборе данных

IV. ЗАКЛЮЧЕНИЕ

Рак молочной железы является серьезной проблемой здравоохранения во всем мире, и раннее выявление с помощью маммографии имеет решающее значение для снижения показателей смертности и заболеваемости. Микрокальцификации являются основным ранним признаком рака молочной железы, и их выявление и классификация могут быть сложной задачей для радиологов. Для повышения точности были предложены методы автоматизированного обнаружения (CADe), и глубокое обучение продемонстрировало большой потенциал в этом отношении. Сегментация изображения является важным шагом в анализе изображений и может помочь в выявлении микрокальцинатов. В этой статье был проведен обзор предыдущих исследований глубоких сверточных нейронных сетей и кластеров микрокальцификации для выявления рака молочной железы и представлен алгоритм предварительной обработки изображений для сегментации и классификации изображений с использованием глубокого обучения. Алгоритмы предварительной обработки, используемые для сегментации изображений и дальнейшей классификации с использованием глубокого обучения, могут помочь улучшить обнаружение и классификацию микрокальцинатов и повысить уровень точности в процессе обнаружения.

V. НАПРАВЛЕНИЯ НА БУДУЩЕЕ

За текущими шагами предварительной обработки последуют:

- Увеличение изображения путем поворота изображений под разными углами обзора из-за количества данных, необходимых при глубоком обучении для обучения модели.
- Сегментация изображений и извлечение признаков с использованием нейронных сетей.
- Классификация изображений маммограммы.

Список литературы

[1] S. Azam et al., “Mammographic microcalcifications and risk of breast cancer,” Br. J. Cancer, vol. 125, pp. 759–765, June 2021.

[2] K. Kumar Singh et al., “Deep Learning Capabilities for the Categorization of Microcalcification,” Int. J. Environ. Res. Public. Health, vol. 19, pp. 2159, February 2022.

[3] A.N.R. Hakim, P. Prajitno, and D.S. Soejoko, “Microcalcification detection in mammography image using computer-aided detection based on convolutional neural network,” AIP Conference Proceedings, vol. 2346, pp. 040001-1 –040007, March 2021. https://aip.scitation.org/doi/10.1063/5.0047828

[4] J. Wang et al., “Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning,” Sci. Rep. Nature Publishing Group, vol. 6, pp. 27327, June 2016.

[5] W. Auccahuasi, C. Delrieux, F. Sernaqué1, E. Flores, and N. Moggiano, “Detection of Microcalcifications in Digital Mammography Images, using Deep Learning Techniques, based on Peruvian Casuistry,” 2019 E-Health and Bioengineering Conference (EHB).Iasi, Romania: IEEE, pp. 1–4, November 2019.

[6] S. O’Grady and M.P. Morgan, “Microcalcifications in breast cancer: From pathophysiology to diagnosis and prognosis,” Biochim. Biophys. Acta BBA - Rev. Cancer, vol. 1869, pp. 310–320. April 2018.

[7] B. Bennani-Baiti and P.A. Baltzer, “MR Imaging for Diagnosis of Malignancy in Mammographic Microcalcifications: A Systematic Review and Meta-Analysis,” Radiology, vol. 283, pp. 692–701, June 2017.

[8] Y. Ouyang et al., “A review of ultrasound detection methods for breast microcalcification,” Math. Biosci. Eng., vol. 16, 1761–1785, 2019.

[9] S. Iranmakani et al., “A review of various modalities in breast imaging: technical aspects and clinical outcomes,” Egypt. J. Radiol. Nucl. Med., vol. 51, pp. 57. April 2020.

[10] S. Ekici and H. Jawzal, “Breast cancer diagnosis using thermography and convolutional neural networks,” Med. Hypotheses, vol. 137, pp. 109542, April 2020.

[11] O. Ginsburg et al., “Breast cancer early detection: A phased approach to implementation,” Cancer, vol. 126 Suppl 10, pp. 2379–2393, May 2020.

[12] M.S. Kahnouei, M. Giti, M.A. Akhaee, and A. Ameri, “Microcalcification Detection in Mammograms Using Deep Learning,” Iran. J. Radiol. Brieflands, vol. 19, 2022.

[13] I.L. Milankovic, N.V. Mijailovic, N.D. Filipovic, and A.S. Peulic, “Acceleration of Image Segmentation Algorithm for (Breast) Mammogram Images Using High-Performance Reconfigurable Dataflow Computers,” Comput. Math. Methods Med., vol. 2017, pp. 7909282, 2017.

[14] S. Chaudhury et al., “Effective Image Processing and Segmentation-Based Machine Learning Techniques for Diagnosis of Breast Cancer,” Comput. Math. Methods Med., vol. 2022, pp. 6841334, 2022.

[15] J. Wang, Y. Yang, “A context-sensitive deep learning approach for microcalcification detection in mammograms,” Pattern Recognition, vol. 78, pp. 12–22, June 2018.

[16] R. Suresh, A.N. Rao, B.E. Reddy, “Detection and classification of normal and abnormal patterns in mammograms using deep neural network,” Concurr. Comput. Pract. Exp., 2019, vol. 31, pp. e5293.

[17] A.S. Becker et al., “Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer,” Invest. Radiol., vol. 52, pp. 434–440, July 2017.

[18] H. Liu et al. “A deep learning model integrating mammography and clinical factors facilitates the malignancy prediction of BI-RADS 4 microcalcifications in breast cancer screening,” Eur. Radiol., vol. 31, pp. 5902–5912, January 2021.

[19] T.M.A. Basile et al., “Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system,” Phys. Medica Eur. J. Med. Phys. Elsevier, vol. 64, pp. 1–9. June 2019.

[20] C. Cui et al., “The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast,” The Cancer Imaging Archive, 2021.