

DOI: <https://doi.org/10.17816/DD627019>

Классификация наличия злокачественных образований на маммограмме с помощью методов глубокого обучения

А.А. Ибрагимов, С.А. Сенотрусова, А.А. Литвинов, А.А. Беляева, Е.Н. Ушаков, Ю.В. Маркин

Институт системного программирования имени В.П. Иванникова, Москва, Россия

АННОТАЦИЯ

Обоснование. Рак молочной железы — заболевание, которое является одной из основных причин смертности от рака у женщин [1]. Регулярный массовый скрининг, проводимый с помощью маммографии, играет важнейшую роль в своевременном выявлении изменений в тканях молочной железы. Однако начальные стадии патологии часто остаются незамеченными, и их сложно диагностировать [2].

Несмотря на эффективность маммографии в снижении смертности от рака молочной железы, ручной анализ изображений может занимать много времени и быть трудозатратным. Поэтому актуальны попытки автоматизировать этот процесс, например, с помощью систем автоматизированной диагностики (CAD) [3]. Однако в последние годы всё больший интерес вызывают решения на основе нейронных сетей, в частности в биологии и медицине [4–6]. Технологические достижения, использующие искусственный интеллект, уже показали свою эффективность в обнаружении патологий [7, 8].

Цель — создание автоматизированного решения для выявления рака молочной железы на маммограммах.

Материалы и методы. Решение реализовано следующим образом: разработан инструмент на основе глубокой нейронной сети, который позволяет по подаваемому на вход изображению получить вероятность наличия злокачественного образования. Для обучения модели был использован объединённый датасет из открытых наборов данных, таких как MIAS, CBIS-DDSM, INbreast, CMMD, KAU-BCMD, VinDr-Mammo [9–14].

Результаты. Модель классификации, основанная на архитектуре EfficientNet-B3, позволяет достичь при тесте на выборке из комбинированного набора данных до 0,695 по метрике площади под ROC-кривой, 0,88 Sensitivity и 0,9 Specificity. Благодаря обучению на изображениях из разных датасетов, отличающихся по качеству данных и региону получения, модель обладает высокой обобщающей способностью, что является ещё одним преимуществом. Кроме того, для повышения эффективности использовались такие методы, как предварительная обрезка изображений и аугментации в процессе обучения.

Заключение. Результаты экспериментов показали, что модель на высоком уровне достоверности может выявлять злокачественные образования. Полученные высокие метрики качества обеспечивают значительный потенциал для внедрения данного метода в автоматизированную диагностику, например, в качестве дополнительного мнения для медицинских специалистов.

Ключевые слова: рак молочной железы; маммография; глубокое обучение; нейронные сети; искусственный интеллект.

Как цитировать:

Ибрагимов А.А., Сенотрусова С.А., Литвинов А.А., Беляева А.А., Ушаков Е.Н., Маркин Ю.В. Классификация наличия злокачественных образований на маммограмме с помощью методов глубокого обучения // Digital Diagnostics. Т. 5, № S1 С. 137–139. DOI: <https://doi.org/10.17816/DD627019>

СПИСОК ЛИТЕРАТУРЫ

1. Milroy M.J. Cancer statistics: Global and national. In: Quality Cancer Care: Survivorship Before, During and After Treatment. Hopewood P., Milroy M.J., editors. Springer, 2018.
2. Mainiero MB, Moy L, Baron P, et al. ACR appropriateness criteria breast cancer screening // Journal of the American College of Radiology. 2017. Vol. 14, N 11S. P. S383–S390. doi: 10.1016/j.jacr.2017.08.044
3. Elter M., Horsch A. CADx of mammographic masses and clustered microcalcifications: a review // Medical physics. 2009. Vol. 36, N 6. P. 2052–2068. doi: 10.1118/1.3121511
4. Kegeles E, Naumov A, Karpulevich EA, Volchkov P, Baranov P. Convolutional neural networks can predict retinal differentiation in retinal organoids // Front. Cell. Neurosci. 2020. Vol. 14. P. 171. doi: 10.3389/fncel.2020.00171
5. Ibragimov A, Senotrusova S, Markova K, et al. Deep semantic segmentation of angiogenesis images // Int. J. Mol. Sci. 2023. Vol. 24, N 2. doi: 10.3390/ijms24021102
6. Naumov A., Ushakov E., Ivanov A., et al. EndoNuke: Nuclei detection dataset for estrogen and progesterone stained IHC endometrium scans // Data (Basel). 2022. Vol. 7, N 6. doi: 10.3390/data7060075

Рукопись получена: 16.02.2024

Рукопись одобрена: 05.03.2024

Опубликована online: 30.06.2024

7. Dembrower K, Wählin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study // *The Lancet Digital Health*. 2020. Vol. 2, N 9. P. e468–e474. doi: 10.1016/S2589-7500(20)30185-0
8. Jiang Y., Edwards A.V., Newstead G.M. Artificial intelligence applied to breast MRI for improved diagnosis // *Radiology*. 2021. Vol. 298, N 1. P. 38–46. doi: 10.1148/radiol.2020200292
9. Suckling J. The mammographic image analysis society digital mammogram database // *Exerpta Medica International Congress*. 1994. Vol. 1069. P. 375–378.
10. Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research // *Sci. Data*. 2017. Vol. 4. P. 170177. doi: 10.1038/sdata.2017.177
11. Moreira IC, Amaral I, Domingues I, et al. INbreast: toward a full-field digital mammographic database // *Acad. Radiol*. 2012. Vol. 19, N 2. P. 236–248. doi: 10.1016/j.acra.2011.09.014
12. Cui C., Li L., Cai H., et al. The Chinese mammography database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast // *Data Cancer Imaging Arch*. 2021. doi: 10.7937/tcia.eqde-4b16
13. Alsolami AS, Shalash W, Alsaggaf W, et al. King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD) // *Data Basel*. 2021. Vol. 6, N 11. P. 111. doi: 10.3390/data6110111
14. Nguyen HT, Nguyen HQ, Pham HH, et al. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography // *Sci. Data*. 2023. Vol. 10, N 1. P. 277. doi: 10.1038/s41597-023-02100-7

DOI: <https://doi.org/10.17816/DD627019>

Classification of the presence of malignant lesions on mammogram using deep learning

Alisher A. Ibragimov, Sofya A. Senotrusova, Arseniy A. Litvinov, Aleksandra A. Beliaeva, Egor N. Ushakov, Yuri V. Markin

Institute for System Programming, Moscow, Russia

ABSTRACT

BACKGROUND: Breast cancer is one of the leading causes of cancer-related mortality in women [1]. Regular mass screening with mammography plays a critical role in the early detection of changes in breast tissue. However, the early stages of pathology often go undetected and are difficult to diagnose [2].

Despite the effectiveness of mammography in reducing breast cancer mortality, manual image analysis can be time consuming and labor intensive. Therefore, attempts to automate this process, for example using computer-aided diagnosis systems, are relevant [3]. In recent years, however, solutions based on neural networks have gained increasing interest, especially in biology and medicine [4–6]. Technological advances using artificial intelligence have already demonstrated their effectiveness in pathology detection [7, 8].

AIM: The study aimed to develop an automated solution to detect breast cancer on mammograms.

MATERIALS AND METHODS: The solution is implemented as follows: a deep neural network-based tool has been developed to obtain the probability of malignancy from the input image. A combined dataset from public datasets such as MIAS, CBIS-DDSM, INbreast, CMMD, KAU-BCMD, and VinDr-Mammo [9–14] was used to train the model.

RESULTS: The classification model, based on the EfficientNet-B3 architecture, achieved an area under the ROC curve of 0.95, a sensitivity of 0.88, and a specificity of 0.9 when tested on a sample from the combined dataset. The model's high generalization ability, which is another advantage, was demonstrated by its ability to perform well on images from different datasets with varying data quality and acquisition regions. Furthermore, techniques such as image pre-cropping and augmentations during training were used to enhance the model's performance.

CONCLUSIONS: The experimental results demonstrated that the model is capable of accurately detecting malignancies with a high degree of confidence. The obtained high-quality metrics offer a significant potential for implementing this method in automated diagnostics, for instance, as an additional opinion for medical specialists.

Keywords: breast cancer; mammography; deep learning; neural networks; artificial intelligence.

To cite this article:

Ibragimov AA, Senotrusova SA, Litvinov AA, Beliaeva AA, Ushakov EN, Markin YuV. Classification of the presence of malignant lesions on mammogram using deep learning. *Digital Diagnostics*. 2024;5(S1):137–139. DOI: <https://doi.org/10.17816/DD627019>

Received: 16.02.2024

Accepted: 05.03.2024

Published online: 30.06.2024

REFERENCES

1. Milroy MJ. Cancer statistics: Global and national. In: *Quality Cancer Care: Survivorship Before, During and After Treatment*. Hopewood P, Milroy MJ, editors. Springer; 2018.
2. Mainiero MB, Moy L, Baron P, et al. ACR appropriateness criteria breast cancer screening. *Journal of the American College of Radiology*. 2017;14(11S):S383–S390. doi: 10.1016/j.jacr.2017.08.044
3. Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: a review. *Medical physics*. 2009;36(6):2052–2068. doi: 10.1118/1.3121511
4. Kegeles E, Naumov A, Karpulevich EA, Volchikov P, Baranov P. Convolutional neural networks can predict retinal differentiation in retinal organoids. *Front. Cell. Neurosci.* 2020;14:171. doi: 10.3389/fncel.2020.00171
5. Ibragimov A, Senotrusova S, Markova K, et al. Deep semantic segmentation of angiogenesis images. *Int. J. Mol. Sci.* 2023;24(2). doi: 10.3390/ijms24021102
6. Naumov A, Ushakov E, Ivanov A, et al. EndoNuke: Nuclei detection dataset for estrogen and progesterone stained IHC endometrium scans. *Data (Basel)*. 2022;7(6). doi: 10.3390/data7060075
7. Dembrower K, Wåhlin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *The Lancet Digital Health*. 2020;2(9):e468–e474. doi: 10.1016/S2589-7500(20)30185-0
8. Jiang Y, Edwards AV, Newstead GM. Artificial intelligence applied to breast MRI for improved diagnosis. *Radiology*. 2021;298(1):38–46. doi: 10.1148/radiol.2020200292
9. Suckling J. The mammographic image analysis society digital mammogram database. *Excerpta Medica International Congress*. 1994;1069:375–378.
10. Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data*. 2017;4:170177. doi: 10.1038/sdata.2017.177
11. Moreira IC, Amaral I, Domingues I, et al. INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* 2012;19(2):236–248. doi: 10.1016/j.acra.2011.09.014
12. Cui C, Li L, Cai H, et al. The Chinese mammography database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast. *Data Cancer Imaging Arch*. 2021. doi: 10.7937/tcia.eqde-4b16
13. Alsolami AS, Shalash W, Alsaggaf W, et al. King Abdulaziz University breast cancer mammogram dataset (KAU-BCMD). *Data Basel*. 2021;6(11):111. doi: 10.3390/data6110111
14. Nguyen HT, Nguyen HQ, Pham HH, et al. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Sci. Data*. 2023;10(1):277. doi: 10.1038/s41597-023-02100-7

ОБ АВТОРАХ

*** Ибрагимов Алишер Ахмад Угли;**

ORCID: 0000-0002-4406-4562;

eLibrary SPIN: 3540-3992;

e-mail: ibragimov@ispras.ru

Сенотрусова Софья Андреевна;

ORCID: 0000-0003-0960-8920;

eLibrary SPIN: 4872-3388;

e-mail: senotrusova@ispras.ru

Литвинов Арсений Александрович;

ORCID: 0009-0000-3561-3817;

e-mail: filashkov@ispras.ru

Беляева Александра Андреевна;

e-mail: belyaeva.a@ispras.ru

Ушаков Егор Николаевич;

ORCID: 0000-0001-8370-6911;

e-mail: ushakov@ispras.ru

Маркин Юрий Витальевич;

ORCID: 0000-0003-1145-5118;

eLibrary SPIN: 8440-9532;

e-mail: ustas@ispras.ru

AUTHORS' INFO

*** Alisher A. Ibragimov;**

ORCID: 0000-0002-4406-4562;

eLibrary SPIN: 3540-3992;

e-mail: ibragimov@ispras.ru

Sofya A. Senotrusova;

ORCID: 0000-0003-0960-8920;

eLibrary SPIN: 4872-3388;

e-mail: senotrusova@ispras.ru

Arsenii A. Litvinov;

ORCID: 0009-0000-3561-3817;

e-mail: filashkov@ispras.ru

Aleksandra A. Belyaeva;

e-mail: belyaeva.a@ispras.ru

Egor N. Ushakov;

ORCID: 0000-0001-8370-6911;

e-mail: ushakov@ispras.ru

Yury V. Markin;

ORCID: 0000-0003-1145-5118;

eLibrary SPIN: 8440-9532;

e-mail: ustas@ispras.ru

* Автор, ответственный за переписку / Corresponding author