DOI: 10.35191/medsoft_2022_1_34_50_54

FULLY INTEGRATED DECISION-SUPPORT SYSTEM FOR DETECTION AND SEGMENTATION OF BREAST LESIONS IN DIGITAL MAMMOGRAM

Robert Hrubý, Daniel Kvak, Anna Chromcová, Marek Biroš

Abstract

Breast cancer is one of the most prevalent forms of cancer affecting women. Detection of suspicious lesions on mammographic images is considered a challenging task due to the variability of lesion sizes and shapes, the problematic margins of the findings, and some extremely small lesions that are difficult to localize. With the increasing availability of digitized clinical archives and the development of complex deep learning (DL) methods, we are witnessing a trend towards the integration of robust computer-aided detection (CAD) systems to assist in the automatic segmentation of lesions on mammograms to aid in the diagnosis of breast cancer. This study presents deep learning–based automatic detection algorithm (DLAD), directly implemented in picture archiving and communication system (PACS) to aid in improving the radiologist's workflow. The proposed DLAD is evaluated on INbreast dataset with a sample size of n=138 (71 [51.45%] BI-RADS 4/5/6 images, 67 [48.55%] BI-RADS 1 images). Preliminary results show a sensitivity of 0.9296 [95% CI 0.8701-0.9891], specificity of 0.7273 [0.6207-0.8339] and IoU of 0.5661, indicating a low false negative rate while maintaining a reasonable false positive rate.

Keywords

Breast Cancer, Breast Lesion Detection, Computer-Aided Detection, Deep Learning, Image Segmentation, Mammogram, Picture Archiving and Communicating System

1 Introduction

Breast cancer is the leading cause of cancer deaths in women worldwide [15]. Nearly 2.3 million new cases were diagnosed and 685,000 deaths were attributed to breast cancer in 2020 [36]. Early detection of suspicious lesions is therefore crucial for successful treatment and reduction of mortality. [5, 24, 28] have highlighted that frequent mammography screening can reduce mortality through early detection before it spreads to other healthy organs and tissues [8]. Regular mammography screening has been successful in reducing breast cancer mortality [30], and adherence to it is high [38]. However, mammograms must be manually analyzed by radiologists determine shapes and types (Figure 1) of any suspicious area in the breast and localize potentially malignant lesions. While this process is considered crucial, it is time-consuming [37] and of strong subjectivity among the evaluating radiologists [17].

[4] http://www.eng.usf.edu/cvprg/mammography/database.html

2 Background

The problem of missed or overlooked lesions still persists, despite modern fullfield digital mammography. [9] reports that the sensitivity and specificity of screening mammography is around 80% and 90%, respectively. [4, 6, 14] observe that double reading improves the performance of mammographic evaluation, showing that there is considerable space for improvement.

2.1 Related Works

The earliest concepts for detecting abnormalities on mammographic images were introduced in the 1960s [3]. Initially, the research and development focused on reducing errors caused by human fatigue or subjectivity. Today, CAD systems can serve two different roles: as a collaborative assistant that directs the radiologist's attention to suspicious areas in the mammogram, or as an independent "reader" that performs an overall assessment of the entire examination without any radiologist intervention.

Deep neural networks, popular paradigms for automated medical image diagnosis, represent the state-of-the-art in computer vision. Models based on deep neural networks have demonstrated robust results in segmentation problems for mammographic images [1, 25, 29], given the large datasets that have become available, such as DDSM [18] or INbreast [26]. Multiple studies, including [2, 7, 11, 19, 23, 27, 33], also addressed the evaluation (mostly retrospective) of existing commercial solutions.

3 Methodology

The aim of this study is to evaluate the performance of DLAD system (Carebot AI MMG v1.0) for breast lesion segmentation when used as decision-support system, compared to the original radiologist decision. As presented on Figure 2, the software is integrated into the standard reading environment (PACS), including a breast density assessment module, but this functionality is not investigated in this study. Proposed DLAD is not a certified medical device.

3.1 Datasets

Mammograms with pixel-level annotations were needed to train the model for lesion segmentation. We utilized the Digital Database for Screening Mammography[4] (Table 1), which contains 10,480 digitized screening mammography images in CC and LMO view positions with pixel-level lesion annotations. Malignant lesions have histological evidence. Because our primary objective is the detection of any lesions on digital mammography, the images used for training include both histologically proven cancerous findings and benign lesions that were recalled for further examination but later found to be benign. To increase the amount of used images, all projections were reoriented to the right-sided position for analysis, and all computations were performed with right-side projections. The model was then evaluated on INbreast dataset with sample size of n=138.

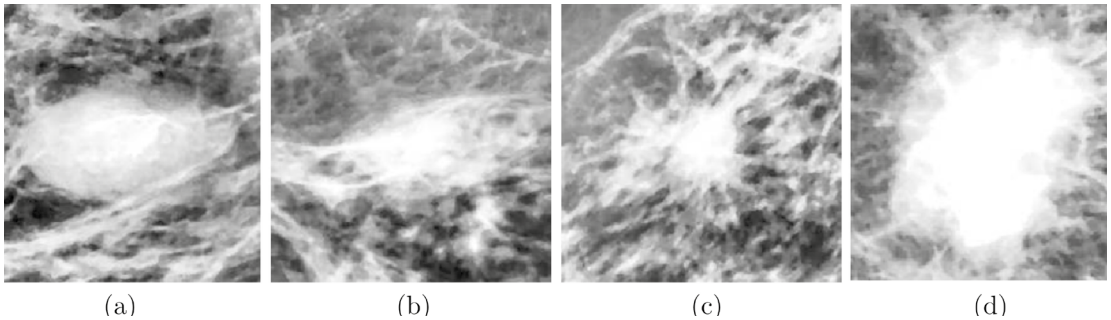

Figure 1 – Examples of various benign (a, b) and malignant (c, d) lesions from digital mammogram.

DOI: 10.35191/medsoft_2022_1_34_50_54

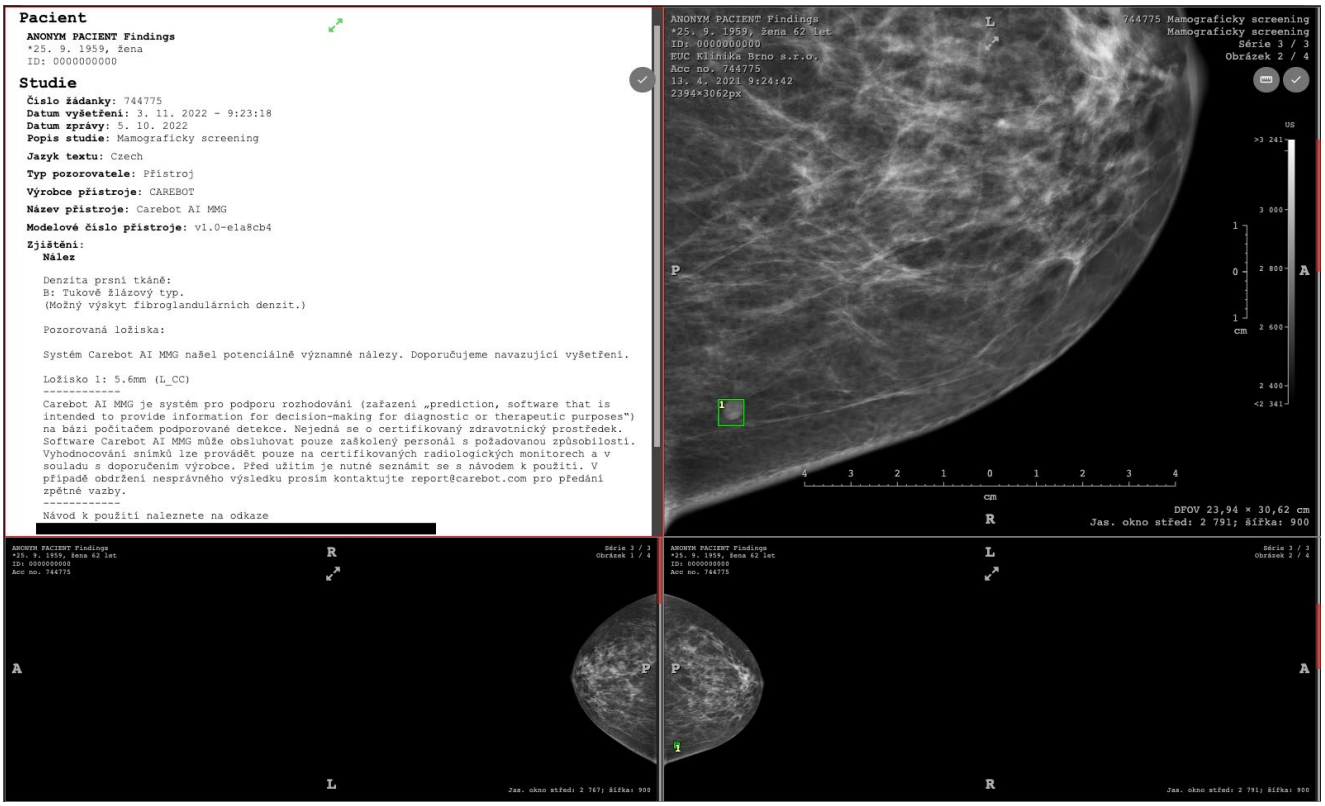

Figure 2 – User interface of DLAD software implemented in PACS (Dicompass Cloud v2.2.8. by Medoro s.r.o.).

Dataset	n	View Position	Format
[18] DDSM	10,480	CC, MLO	DICOM
[26] INbreast	410	CC, MLO	DICOM

Table 1 – Detailed information of the used DDSM and INbreast mammography datasets.

3.2 Data Preprocessing

Suspicious lesions can be observed on mammograms as high-density areas. As shown in Figure 3, original (before preprocessing) mammograms indicate poor visualization of the dense mammary gland and breast periphery. Proposed model utilizes contrast limited adaptive histogram equalization (CLAHE): a technique that limits the histogram equalization amplification by clipping the histogram to a user-defined value (clip limit) [32]. Application of CLAHE method and segmentation of the non-relevant metadata areas (after preprocessing) improve image quality with emphasis on the nipple, areola, skin, subcutaneous fat and some peripheral Cooper's ligaments [10].

3.3 Model Architecture

Since Ronneberger et al. initially proposed the U-Net architecture in [31], it has become the state-of-the-art technique for biomedical applications, being utilized in semantic segmentation, object detection, and more. The basic building blocks consist of a downsampling and an upsampling path. These two branches form a U-shape, as shown in Figure 4. The proposed architecture introduced the skip connections which added a significant advantage compared to the predecessors. Specifically, it helps to recover spatial information that is lost during the downsampling steps due to the pooling operations.

The encoder block (also known as *backbone*) is usually a sequence of convolutional neural networks (CNN) used for classification tasks, but missing the final dense layers [35]. These

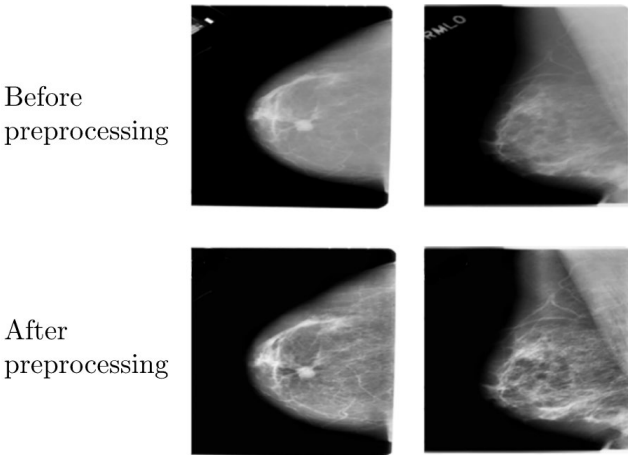

Figure 3 – Mammographic images before and after preprocessing

types of networks are called Fully Convolutional Network (FCN). In this study, VGG-16 pre-trained on the ImageNet dataset [13] was used as a backbone model.

3.4 Implementation

The DLAD software leverages the DICOMweb™ protocol to communicate with a connected PACS. DICOMweb™ is a DICOM communication standard for webbased medical imaging that allows web application developers to transfer medical data between AEs using proprietary tools. It is a set of services using the MIME type *multipart/related* (HTTP) Internet protocol in a RESTful interface [12]. The basic services implemented by the DICOMweb™ standard can be defined as commands to store and query medical images and related information. The standard is formally defined in the DICOM PS3.18 *Web Services* document [16].

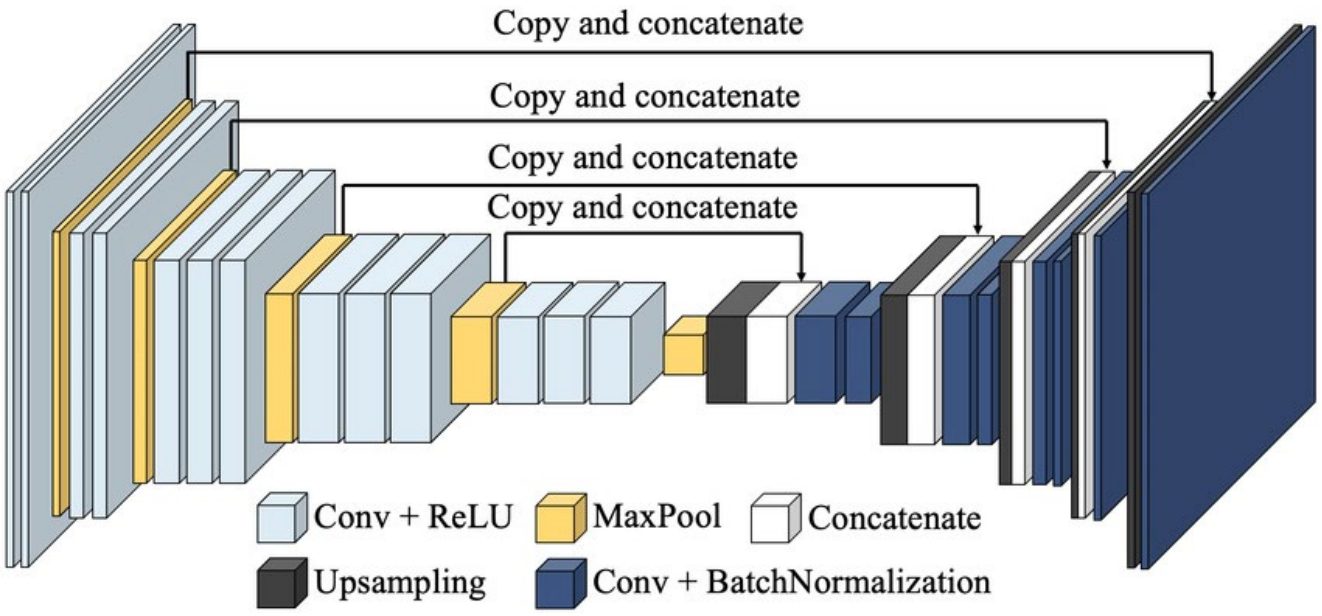

Figure 3 – VGG U-Net model architecture [34]

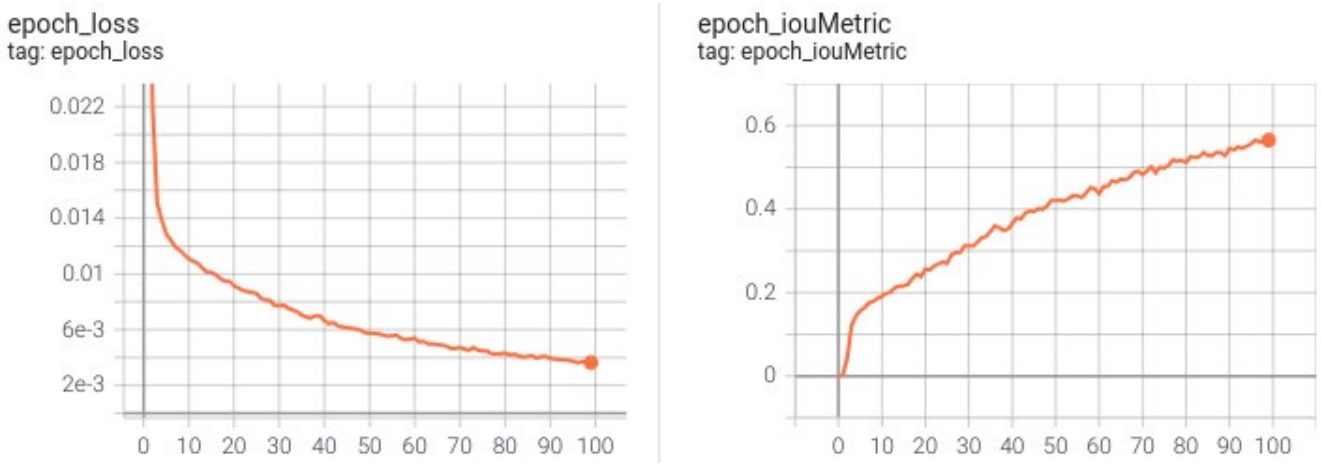

Figure 5 – TensorBoard visualization of loss and IoU metric during DLAD model training

3.5 Statistical Analysis

Performance of the proposed system was quantified by means of accuracy *(Acc)*, sensitivity *(Se)*, specificity *(Sp)*, positive *(PPV)* and negative predictive value *(NPV)* and Intersection over Union *(IoU)*. Formulas can be found in Table 2. All confidence intervals *(CI)* were constructed at a two-tailed 95% confidence level.

Acc	0.8321 [0.7697-0.8945]	*Acc = (TP + TN) / (P + N)*
Se	0.9296 [0.8701-0.9891]	*Se = TP / (TP + FN)*
Sp	0.7273 [0.6207-0.8339]	*Sp = TN / (FP + TN)*
PPV	0.7857 [0.6903-0.8811]	*PPV = TP / (TP + FP)*
NPV	0.9057 [0.8357-0.9757]	*NPV = TN / (TN + FN)*
IoU	0.5661	*IoU = area of overlap / area of union*

Table 2 – Performance of the DLAD system on a test sample from the INbreast dataset. The performance was quantified using accuracy, sensitivity, specificity, positive and negative predictive value and intersection over union.

4 Results

A total of 138 individual images (71 [51.45%] BI-RADS 4/5/6 images, 67 [48.55%] BI-RADS 1 images) with determined ground truth and pixel-level annotations were analyzed. The proposed DLAD system correctly classified 114 out of 138 test images (*Acc*: 0.8321 [95% CI 0.7697-0.8945]). 18 BI-RADS 1 mammograms were misclassified (FP) as containing a lesion (*Sp*: 0.7273 [0.6207-0.8339]). A higher false positive rate was expected outcome because the DL algorithm was trained to identify even benign findings as abnormal. Additional 5 BI-RADS 4/5/6 images were misclassified (FN) as BI-RADS 1 despite including one or more lesions (*Se*: 0.9296 [0.8701-0.9891]).

The PPV, i.e. the probability of a positive finding if an image is labeled as BI-RADS 4/5/6, was 0.7857 [0.6903-0.8811] for DLAD. The NPV, i.e. the probability that a patient is without any finding when the image was classified as BI-RADS 1, was 0.9057 [0.8357-0.9757].

5 Discussion

Although we were able to achieve promising sensitivity and reasonable specificity on a small sample size, current publications of commercially applicable solutions are working with larger

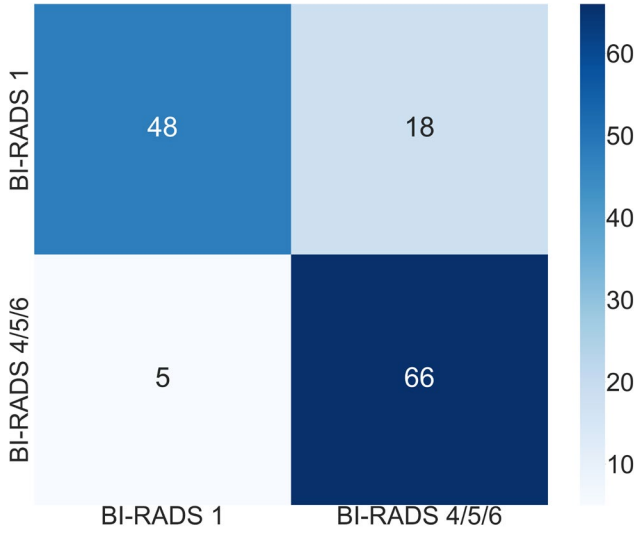

Figure 6 – Confusion matrix demonstrating the results of the DLAD software on test data sampled from INbreast dataset.

volumes of test data. [19] (n=1,238) verified own CAD system and achieved Se and Sp of 0.76. and 0.88, respectively. [33] (n=8,805) inspected three commercially available solutions and achieved Se of 0.82 for AI-1, 0.67 for AI-2, 0.67 for AI-3 and 0.77 for first--reader radiologist, and 0.80 for second-reader radiologist. [7] (n=2,396) used a deep learning-based system on retrospectively assessed data from a mammography screening program in north-western Germany: the inspected algorithm was able to detect and correctly localise 27.5% [95% CI: 23.3–32.3%], and 12.2% [95% CI: 9.5–15.5%] of the FN and MS cases on the prior mammogram, respectively. [27] (n=240) analyzed a comparably large test cohort as in our study but instead focused on improved diagnostic performance when using AI: the average AUC across readers was 0.769 [95% CI: 0.724-0.814] without AI and 0.797 [95% CI: 0.754-0.840] with AI. Average sensitivity was increased by 0.033 when using AI support [p=0.021].

Some limitations of our test may include the exclusion of BI--RADS 3 category images from the evaluation. [20–22] suggest that this particular class is often problematic to assess, both for reading by the radiologist and for AI assessment. However, the question of the potential added benefits of AI in mammography screening remains mostly unanswered. [23] examines benefits when used either as a standalone system or within a decision-referral approach, compared with the original radiologist decision. The exemplary configuration of the AI system in standalone mode achieved a Se of 0.842 (95% CI 0.824–0.858) and a Sp of 0.895 (0.890–0.899) on internal-test data, and a Se of 0.846 (0.833–0.859) and a Sp of 0.913 (0.911–0.915) on external-test data, but was less accurate than the average unaided radiologist. By contrast, the simulated decision-referral approach significantly improved upon radiologist Se by 2.6% and Sp by 1%, corresponding to a triaging performance at 0.63 on the external dataset; the AUROC was 0.982 (95% CI 0.978–0.986) on the subset of studies assessed by AI, surpassing radiologist performance. Automatic triage using commercially available software was also addressed in [11] (n=7,364) When including 60%, 70%, or 80% of women with the lowest AI scores in the no radiologist stream, the proportion of screendetected cancers that would have been missed were 0%, 0.3% (95% CI 0.0–4.3), or 2.6% (1.1–5.4), respectively.

6 Conclusion

In recent years, a continuous increase in mammography examinations has been observed globally. Along with the increasing amount, there is also increasing demand on the assessed radiologists to streamline the detection and localization of potentially suspicious areas. The SotA technological level allows the introduction of AI CAD solutions in the analysis process. This can significantly help medical staff both in time savings and indicating unclear or difficult-to-recognise cases, thereby improving the accuracy of diagnosis and the efficiency of examinations.

The aim of this study was the introduction, implementation and initial evaluationof DLAD software, which utilizes DL algorithms to analyse mammogramsand automatically detect suspicious lesions. In the context of a clinical practice,the software should aim to provide an additional expert reading capability to alert the physician to areas that he might overlooked or with which the doctor is unsure.

We trained the DL algorithm on a publicly available DDSM dataset with pixel-level annotations and implemented it in PACS using DICOMweb™ services. The software was subsequently evaluated on INbreast dataset with a sample size of n=138, achieving a sensitivity of 0.9296 [95% CI 0.8701-0.9891], specificity of 0.7273 [0.6207-0.8339] and IoU of 0.5661, indicating a low false negative rate while maintaining a reasonable false positive rate.

References

[1.] Abdelhafiz, D., Bi, J., Ammar, R., Yang, C. & Nabavi, S. Convolutional neural network for automated mass segmentation in mammography. BMC Bioinformatics. **21**, 1-19 (2020)

[2.] Abdelrahman, L., Al Ghamdi, M., Collado-Mesa, F. & Abdel-Mottaleb, M. Convolutional neural networks for breast cancer detection in mammography: A survey. Computers In Biology And Medicine. **131** pp. 104248 (2021)

[3.] Anaya-Isaza, A., Mera-Jim´enez, L., Cabrera-Chavarro, J., Guachi--Guachi, L., Peluffo-Ord´o˜nez, D. & Rios-Pati˜no, J. Comparison of Current Deep Convolutional Neural Networks for the Segmentation of Breast Masses in Mammograms. IEEE Access. **9** pp. 152206-152225 (2021)

[4.] Anderson, E., Muir, B., Walsh, J. & Kirkpatrick, A. The efficacy of double reading mammograms in breast screening. Clinical Radiology. **49,** 248-251 (1994)

[5.] Broeders, M., Moss, S., Nystr¨om, L., Njor, S., Jonsson, H., Paap, E., Massat, N., Duffy, S., Lynge, E. & Paci, E. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. Journal Of Medical Screening. **19**, 14-25 (2012)

[6.] Brown, J., Bryan, S. & Warren, R. Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. BMj. **312**, 809-812 (1996)

[7.] Byng, D., Strauch, B., Gnas, L., Leibig, C., Stephan, O., Bunk, S. & Hecht, G. AI-based prevention of interval cancers in a national mammography screening program. European Journal Of Radiology. **152** pp. 110321 (2022)

[8.] Coleman, C. Early detection and screening for breast cancer. Seminars In Oncology Nursing. **33,** 141-155 (2017)

[9.] D'Orsi, C., Bassett, L., Feig, S. & Others Breast imaging reporting and data system (BI-RADS). Breast Imaging Atlas, 4th Edn. American College Of Radiology, Reston. (2018)

[10.] Dabass, J., Arora, S., Vig, R. & Hanmandlu, M. Mammogram image enhancement using entropy and CLAHE based intuitionistic fuzzy method. 2019 6th International Conference On Signal Processing And Integrated Networks (SPIN). pp. 24-29 (2019)

[11.] Dembrower, K., Waahlin, E., Liu, Y., Salim, M., Smith, K., Lindholm, P., Eklund, M. & Strand, F. 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. **2**, e468-e474 (2020)

[12.] Demirer, M., Candemir, S., Bigelow, M., Yu, S., Gupta, V., Prevedello, L., White, R., Yu, J., Grimmer, R., Wels, M. & Others A user interface for optimizing radiologist engagement in image data curation for artificial intelligence. Radiology. Artificial Intelligence. **1** (2019)

[13.] Deng, J., Dong, W., Socher, R., Li, L., Li, K. & Fei-Fei, L. Imagenet: A largescale hierarchical image database. 2009 IEEE Conference On Computer Vision And Pattern Recognition. pp. 248-255 (2009)

[14.] Dinnes, J., Moss, S., Melia, J., Blanks, R., Song, F. & Kleijnen, J. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. The Breast. **10**, 455-463 (2001)

[15.] Ferlay, J., H´ery, C., Autier, P. & Sankaranarayanan, R. Global burden of breast cancer. Breast Cancer Epidemiology. pp. 1-19 (2010)

[16.] Genereaux, B., Dennison, D., Ho, K., Horn, R., Silver, E., O'Donnell, K. & Kahn, C. DICOMweb™: Background and application of the web standard for medical imaging. Journal Of Digital Imaging. **31**, 321-326 (2018)

[17.] Giess, C., Wang, A., Ip, I., Lacson, R., Pourjabbar, S. & Khorasani, R. Patient, radiologist, and examination characteristics affecting screening mammography recall rates in a large academic practice. Journal Of The American College Of Radiology. **16**, 411-418 (2019)

[18.] Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K. & Munishkumaran, S. Current status of the digital database for screening mammography. Digital Mammography. pp. 457-460 (1998)

[19.] Kim, E., Kim, H., Han, K., Kang, B., Sohn, Y., Woo, O. & Lee, C. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. Scientific Reports. **8**, 1-8 (2018)

[20.] Lacson, R., Wang, A., Cochon, L., Giess, C., Desai, S., Eappen, S. & Khorasani, R. Factors associated with optimal follow-up in women with BI-RADS 3 breast findings. Journal Of The American College Of Radiology. **17**, 469-474 (2020)

[21.] Lee, K., Talati, N., Oudsema, R., Steinberger, S. & Margolies, L. BI--RADS 3: current and future use of probably benign. Current Radiology Reports. **6**, 1-15 (2018)

[22.] Lee, C., Berg, J. & Berg, W. Cancer yield exceeds 2% for BI-RADS 3 probably benign findings in women older than 60 years in the National Mammography Database. Radiology. **299**, 550-558 (2021)

[23.] Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K. & Umutlu, L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. The Lancet Digital Health. **4**, e507-e519 (2022)

[24.] Meisel, S., Pashayan, N., Rahman, B., Side, L., Fraser, L., Gessler, S., Lanceley, A. & Wardle, J. Adjusting the frequency of mammography screening on the basis of genetic risk: attitudes among women in the UK. The Breast. **24**, 237-241 (2015)

[25.] Mordang, J., Janssen, T., Bria, A., Kooi, T., Gubern-M´erida, A. & Karssemeijer, N. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. International Workshop On Breast Imaging. pp. 35-42 (2016)

[26.] Moreira, I., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. & Cardoso, J. Inbreast: toward a full-field digital mammographic database. Academic Radiology. **19**, 236-248 (2012)

[27.] 27. Pacil`e, S., Lopez, J., Chone, P., Bertinotti, T., Grouin, J. & Fillard, P. Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. Radiology: Artificial Intelligence. **2** (2020)

[28.] Ponti, A., Anttila, A., Ronco, G., Senore, C. & Others Cancer screening in the European Union (2017). Report on the implementation of the council recommendation on cancer screening.. Cancer Screening In The European Union (2017). Report On The Implementation Of The Council Recommendation On Cancer Screening.(2017)

[29.] Ribli, D., Horv´ath, A., Unger, Z., Pollner, P. & Csabai, I. Detecting and classifying lesions in mammograms with deep learning. Scientific Reports. **8**, 1-7 (2018)

[30.] Roder, D., Houssami, N., Farshid, G., Gill, G., Luke, C., Downey, P., Beckmann, K., Iosifidis, P., Grieve, L. & Williamson, L. Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia. Breast Cancer Research And Treatment. **108**, 409-416 (2008)

[31.] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference On Medical Image Computing And Computer-assisted Intervention. pp. 234-241 (2015)

[32.] Sajeev, S., Bajger, M. & Lee, G. Segmentation of breast masses in local dense background using adaptive clip limit-CLAHE. 2015 International Conference On Digital Image Computing: Techniques And Applications (DICTA). pp. 1-8 (2015)

[33.] Salim, M., Waahlin, E., Dembrower, K., Azavedo, E., Foukakis, T., Liu, Y., Smith, K., Eklund, M. & Strand, F. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. JAMA Oncology. **6**, 1581-1588 (2020)

[34.] Shi, J., Dang, J., Cui, M., Zuo, R., Shimizu, K., Tsunoda, A. & Suzuki, Y. Improvement of damage segmentation based on pixel-level data balance using vgg-unet. Applied Sciences. **11**, 518 (2021)

[35.] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:1409.1556. (2014)

[36.] 36. Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal For Clinicians. **71**, 209-249 (2021)

[37.] Wang, L. Early diagnosis of breast cancer. Sensors. **17**, 1572 (2017)

[38.] Weiss, N. Breast cancer mortality in relation to clinical breast examination and breast self-examination. The Breast Journal. **9** pp. S86-S89 (2003)

Contact

Robert Hrubý, BSc.
Carebot s.r.o.
Vyšehradská 430/41
128 00 Praha 2
robert.hruby@carebot.com

Mgr. Daniel Kvak
Carebot s.r.o.
Vyšehradská 430/41
128 00 Praha 2
daniel.kvak@carebot.com

MUDr. Anna Chromcová
Carebot s.r.o.
Vyšehradská 430/41
128 00 Praha 2
anna.chromcova@carebot.com

Mgr. Marek Biroš
Carebot s.r.o.
Vyšehradská 430/41
128 00 Praha 2
marek.biros@carebot.com