

Article

Breast Invasive Ductal Carcinoma Classification on Whole Slide Images with Weakly-Supervised and Transfer Learning

Fahdi Kanavati and Masayuki Tsuneki * 

Medmain Research, Medmain Inc., Fukuoka 810-0042, Japan; fkanavati@medmain.com

* Correspondence: tsuneki@medmain.com; Tel.: +81-92-707-1977

Simple Summary: In this study, we have trained deep learning models using transfer learning and weakly-supervised learning for the classification of breast invasive ductal carcinoma (IDC) in whole slide images (WSIs). We evaluated the models on four test sets: one biopsy ($n = 522$) and three surgical ($n = 1129$) achieving AUCs in the range 0.95 to 0.99. We have also compared the trained models to existing pre-trained models on different organs for adenocarcinoma classification and they have achieved lower AUC performances in the range 0.66 to 0.89 despite adenocarcinoma exhibiting some structural similarity to IDC. Therefore, performing fine-tuning on the breast IDC training set was beneficial for improving performance. The results demonstrate the potential use of such models to aid pathologists in clinical practice.

Abstract: Invasive ductal carcinoma (IDC) is the most common form of breast cancer. For the non-operative diagnosis of breast carcinoma, core needle biopsy has been widely used in recent years for the evaluation of histopathological features, as it can provide a definitive diagnosis between IDC and benign lesion (e.g., fibroadenoma), and it is cost effective. Due to its widespread use, it could potentially benefit from the use of AI-based tools to aid pathologists in their pathological diagnosis workflows. In this paper, we trained invasive ductal carcinoma (IDC) whole slide image (WSI) classification models using transfer learning and weakly-supervised learning. We evaluated the models on a core needle biopsy ($n = 522$) test set as well as three surgical test sets ($n = 1129$) obtaining ROC AUCs in the range of 0.95–0.98. The promising results demonstrate the potential of applying such models as diagnostic aid tools for pathologists in clinical practice.

Keywords: breast; invasive ductal carcinoma; deep learning; weakly-supervised learning; transfer learning; whole slide image



Citation: Kanavati, F.; Tsuneki, M. Breast Invasive Ductal Carcinoma Classification on Whole Slide Images with Weakly-Supervised and Transfer Learning. *Cancers* **2021**, *13*, 5368. <https://doi.org/10.3390/cancers13215368>

Academic Editors: Anke Meyer-Baese, Max Zimmermann, Andreas Stadlbauer

Received: 30 September 2021

Accepted: 23 October 2021

Published: 26 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is one of the leading causes of global cancer incidence [1]. In 2020, there were 2,261,419 new cases (11.7% of all cancer cases) and 684,996 deaths (6.9% of all cancer related deaths) due to breast cancer. Among women, breast cancer accounts for one in four cancer cases and for one in six cancer deaths in the vast majority of countries (159 of 185 countries) [1].

Invasive ductal carcinoma (IDC) (or invasive carcinoma of no special type: ductal NST) is a heterogeneous group of tumors that fail to exhibit sufficient characteristics to achieve classification as a specific histopathological type. Microscopically, there are a wide variety of histopathological characteristics in IDCs. IDC grows in diffuse-sheets, well-defined nests, cords, or as individual (single) cells. Tubular differentiation tends to be well developed, barely detectable, or altogether absent.

Core needle biopsy is frequently used for the management of non-palpable mammogram abnormalities, as it is cost effective and provides an alternative to short-interval follow-up mammography. It is also generally favored over fine-needle aspiration biopsy (FNAB) for the non-operative diagnosis of breast carcinoma, and it could replace open

breast biopsy provided that the quality assurance is acceptable [2,3]. Core needle biopsy allows the evaluation of histopathological features, making it possible to provide a definitive diagnosis of IDC and benign lesions (e.g., fibroadenoma) in over 90% of cases [4]. All these factors highlight the benefit of establishing a histopathological screening system based on core needle biopsy specimens for breast IDC patients. Glass slides of biopsy specimens can be digitised as whole slide images (WSIs) and could benefit from the application of computational histopathology algorithms to aid pathologists as part of a screening system.

Deep learning has found a wide array of applications in computational histopathology in the past few years. The applications from cancer cells classification and segmentation and patient outcome predictions for a variety of organs and diseases [5–18]. Machine learning has been previously applied to various applications of breast histopathology classification [19–24].

In this paper, we trained a WSI breast IDC classification model using transfer learning from ImageNet and weakly-supervised learning. We have also evaluated on the test sets, without fine-tuning, models that had been previously trained on other organs for the classification of carcinomas.

2. Methods

2.1. Clinical Cases and Pathological Records

This is a retrospective study. A total of 2183 H and E (hematoxylin and eosin) stained histopathological specimens of human breast IDC and benign lesions—1154 core needle biopsy and 1028 surgical—were collected from the surgical pathology files of three hospitals: International University of Health and Welfare, Mita Hospital (Tokyo) and Kamachi Group Hospitals (consist of Shinkomonji and Shinkuki hospitals) (Fukuoka) after histopathological review of those specimens by surgical pathologists. The test cases were selected randomly, so the obtained ratios reflected a real clinical scenario as much as possible. All WSIs were scanned at a magnification of x20 using the same Leica Aperio AT2 scanner and were saved SVS file format with JPEG2000.

In addition, we collected 100 WSIs from The Cancer Genome Atlas (TCGA); however, only four benign cases were available.

2.2. Dataset

The pathologists excluded cases that were inappropriate or of poor scanned quality prior to this study. The diagnosis of each WSI was verified by at least two pathologists. Table 1 breaks down the distribution of dataset into training, validation, and test sets. Hospitals that provided histopathological cases were anonymised (e.g., Hospital 1–2). The training set was solely composed of WSIs of core needle biopsy specimens. The test sets were composed of WSIs of core needle biopsy or surgical specimens. The patients' pathological records were used to extract the WSIs' pathological diagnoses and to assign WSI labels. Out of the 191 WSIs with IDC, 96 WSIs were loosely annotated by pathologists. There were about seven annotations per WSI on average. We did not annotate on the carcinoma in situ areas, and some parts of the adjacent stromal area were included in the annotations in order to provide contextual information.

The rest of IDC and benign WSIs were not annotated and the training algorithm only used the WSI labels. Each WSI diagnosis was observed by at least two pathologists, with the final checking and verification performed by a senior pathologist. The senior pathologist only reviewed discordant cases between the two initial pathologists.

2.3. Deep Learning Models

We trained all the models using the partial fine-tuning approach [25]. This method consists of using the weights of an existing pre-trained model and only fine-tuning the affine parameters of the batch normalisation layers and the final classification layer. We have used the EfficientNetB1 architecture [26], as well as B3, with modified input sizes of

224 × 224 px and 512 × 512 px, starting with pre-trained weights from ImageNet. The total number of trainable parameters for EfficientNetB1 was only 63,329.

Table 1. Distribution of WSIs in the different sets.

Set	Source	IDC	Benign	Total
Test	Hospital 1 (biopsy)	289	233	522
	Hospital 1 (surgical)	305	240	545
	Hospital 2 (surgical)	247	237	484
	TCGA (surgical)	96	4	100
Training	Hospital 1 (biopsy)	82	343	425
	Hospital 2 (biopsy)	107	40	147
Validation	Hospital 1 (biopsy)	30	30	60

The training method that we have used in this study is exactly the same as reported in a previous study [27] with the main difference being the use of a partial fine-tuning method. For completeness, we repeat the method here.

To apply the model on the WSIs for training and inference, we performed slide tiling by extracting fixed-sized tiles from tissue regions. We detected the tissue regions by performing a thresholding on a grayscale version of the WSI using Otsu’s method [28], which allows the elimination of most of the white background. During inference, we performed the slide tiling in a sliding window fashion on the tissue regions, using a fixed-size stride that was half the size of the tile. During training, we initially performed random balanced sampling of tiles from the tissue regions, where we maintained an equal balance of positive and negative labelled tiles in the training batch. To do so, we placed the WSIs in a shuffled queue with oversampling of the positive labels to ensure that all the WSIs were seen at least once during each epoch, and we looped over the labels in succession (i.e., we alternated between picking a WSI with a positive label and a negative label). Once a WSI was selected, we randomly sampled $\frac{\text{batch size}}{2}$ tiles from each WSI to form a balanced batch. We then switched into hard mining of tiles. To perform the hard mining, we alternated between training and inference. During inference, the CNN was applied in a sliding window fashion on all of the tissue regions in the WSI, and we then selected the k tiles with the highest probability of being positive. If the tile is from a negative WSI, this step effectively selects the false positives. The selected tiles were placed in a training subset, and once the subset size reached N tiles, a training pass was triggered. We used $k = 4$, $N = 256$, and a batch size of 32.

A subset of WSIs with IDC were loosely annotated ($n = 96$) while the rest had WSI-level labels only ($n = 95$). From the loosely annotated WSIs, we only sampled tiles from the annotated tissue regions. Otherwise, we freely sampled tiles from the entire tissue region.

The models were trained on WSIs at ×10 and ×20 magnifications. We used two input tile sizes: 512 × 512 px and 224 × 224 px. The strides were half the tile sizes. The WSI prediction was obtained by taking the maximum probability from all of the tiles.

We trained the models with the Adam optimisation algorithm [29] with the following parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$. We used a learning rate of 0.001. We applied a learning rate decay of 0.95 every 2 epochs. We used the binary cross entropy loss. We used early stopping by tracking the performance of the model on a validation set, and training was stopped automatically when there was no further improvement on the validation loss for 10 epochs. The model with the lowest validation loss was chosen as the final model.

2.4. Software and Statistical Analysis

The deep learning models were implemented and trained using TensorFlow [30]. AUCs were calculated in python using the scikit-learn package [31] and plotted using matplotlib [32]. The 95% CIs of the AUCs were estimated using the bootstrap method [33] with 1000 iterations.

3. Results

A Deep Learning Model for WSI Breast IDC Classification

The purpose of this study was to train a deep learning model to classify breast IDC in WSIs. We had a total of 1154 biopsy WSIs of which we used 632 for training and 522 for testing. In addition, we used 1129 surgical WSIs obtained from three sources as part of supplementary test sets. We used a transfer learning (TL) approach based on partial fine-tuning [25] to train the models. Figure 1 shows an overview of our training method. We then evaluated the trained models on four tests sets: one biopsy test set and three surgical test sets. We refer to the trained models as TL <magnification> <tile size> <model size>, based on the different configurations.

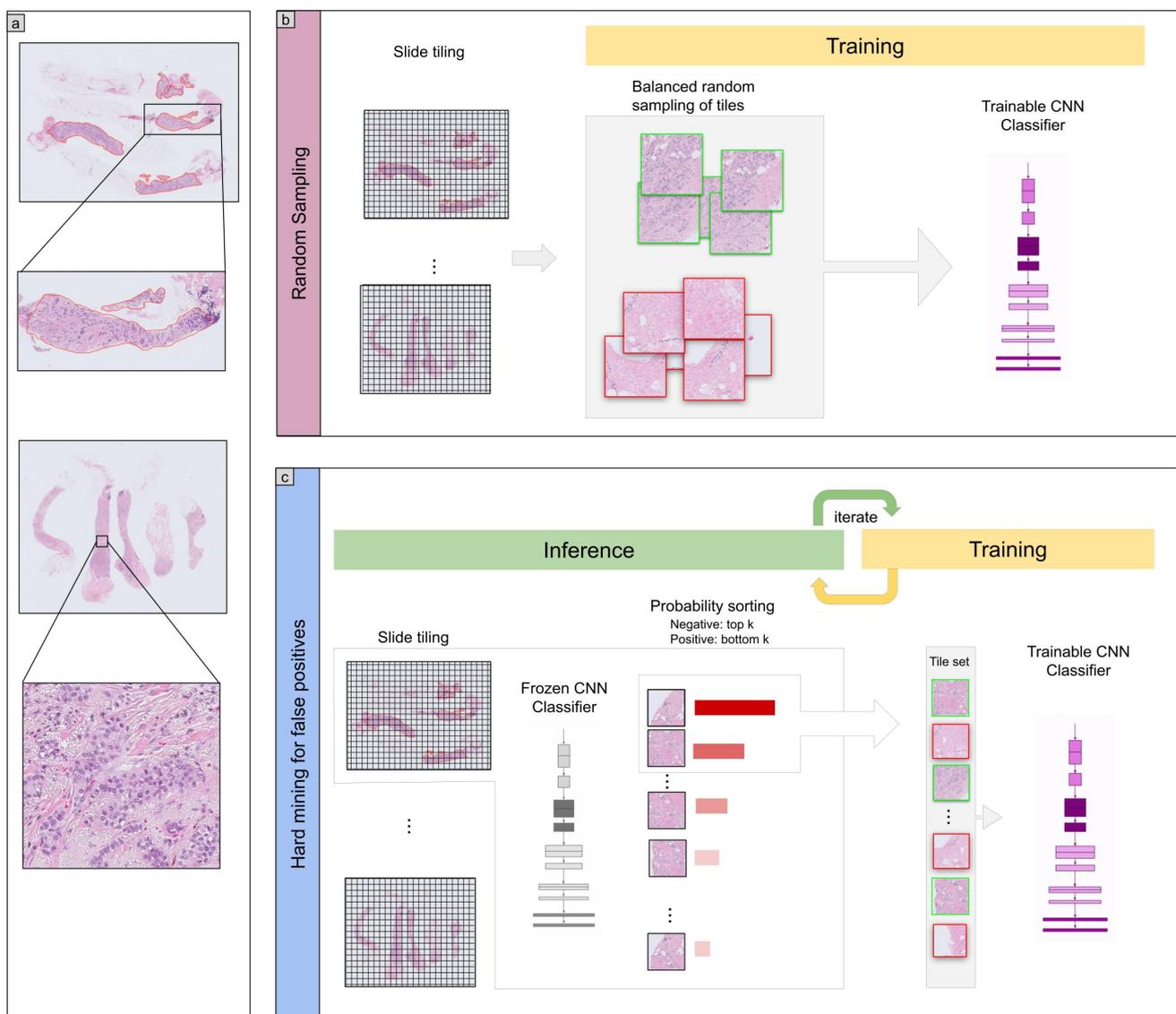


Figure 1. Overview. (a) shows and representative examples of WSIs with a zoom in on the tissue structure. Training consisted of two stages: random sampling and hard mining. In the first stage (b) we randomly sampled tiles from the positive and negative WSIs, restricting the sampling from WSIs that had annotations if they were positive. In the second stage (c) We iteratively alternated between inference and training, relying only on the WSI label. During inference, the model weights were frozen, and it was applied in a sliding window fashion on each WSI. The top k tiles with the highest probabilities were then selected from each WSI. During training the selected tiles were then used to train the model.

As we had at our disposal six models [18,27,34–37] that had been trained specifically on specimens from different organs (stomach, colon, lung, and pancreas), we evaluated those models without fine-tuning on the test sets to investigate whether morphological cancer similarities transfer across organs without additional training.

Table 1 breaks down the distribution of the WSIs in each test set. For each test set, we computed the ROC AUC and log loss, and we have summarised the results in Table 2 and Figures 2 and 3. Figures 4–7 show representative heatmap prediction outputs for true positive, false positive, and false negative. Table 3 shows a confusion matrix breakdown by subtype for the false positives and true negatives using a probability threshold of 0.5. All 10 false positive WSIs were fibroadenomas. Figure 8 shows an overview of representative fibroadenoma histopathology of 10 cases (WSIs) that were falsely predicted as IDC. There were representative histopathologic changes (e.g., proliferative epithelial changes, fibrocystic epithelial changes, and stromal changes) [38] in falsely predicted fibroadenomas (Figure 8); the proliferative findings could be the potential cause of the false positive.

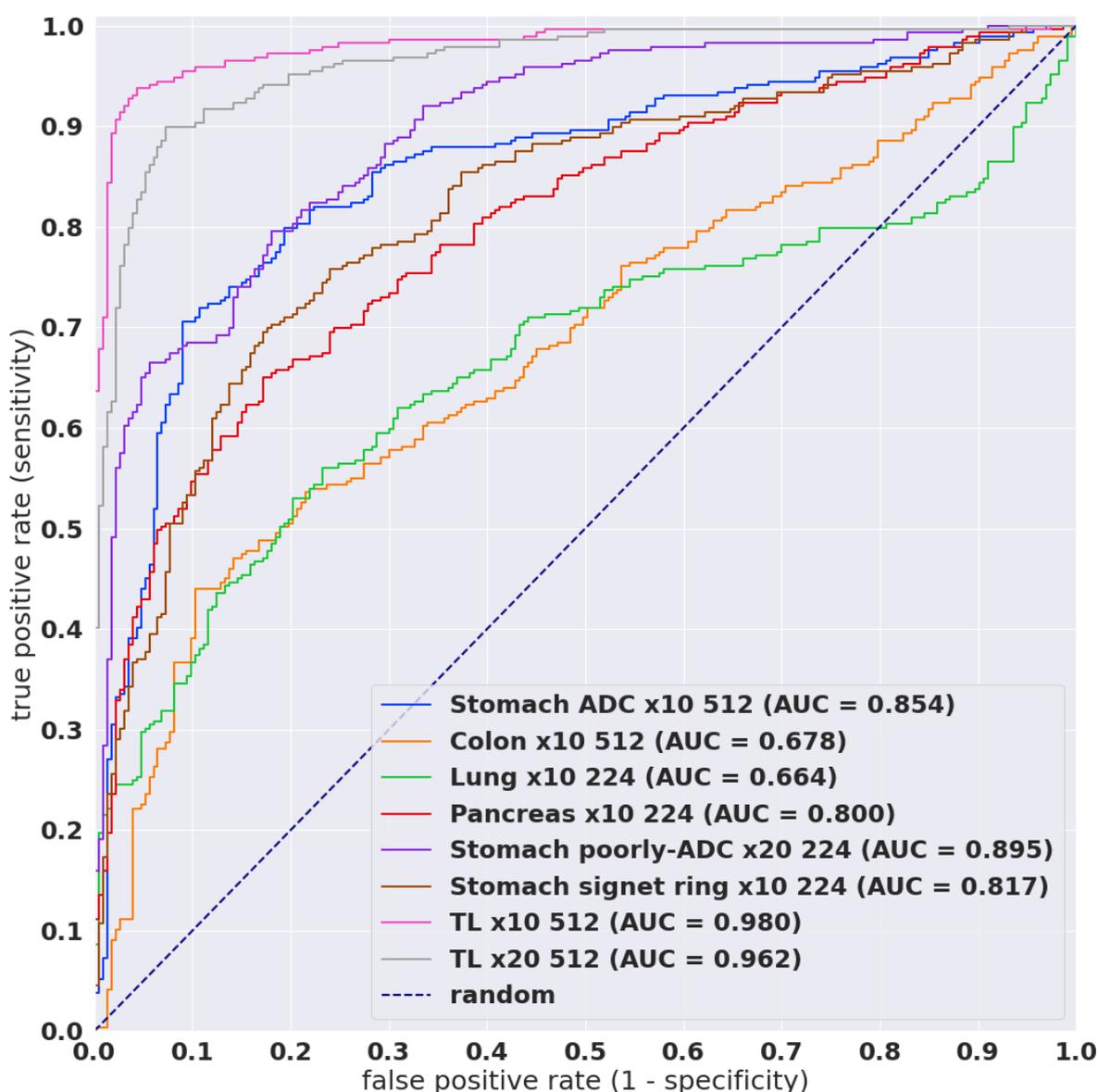


Figure 2. ROC curves for the various existing models as well as models trained via transfer learning (TL) on core needle biopsy test set from Hospital 1. The trained models were TL x10 512 B1 and TL x20 512 B1.

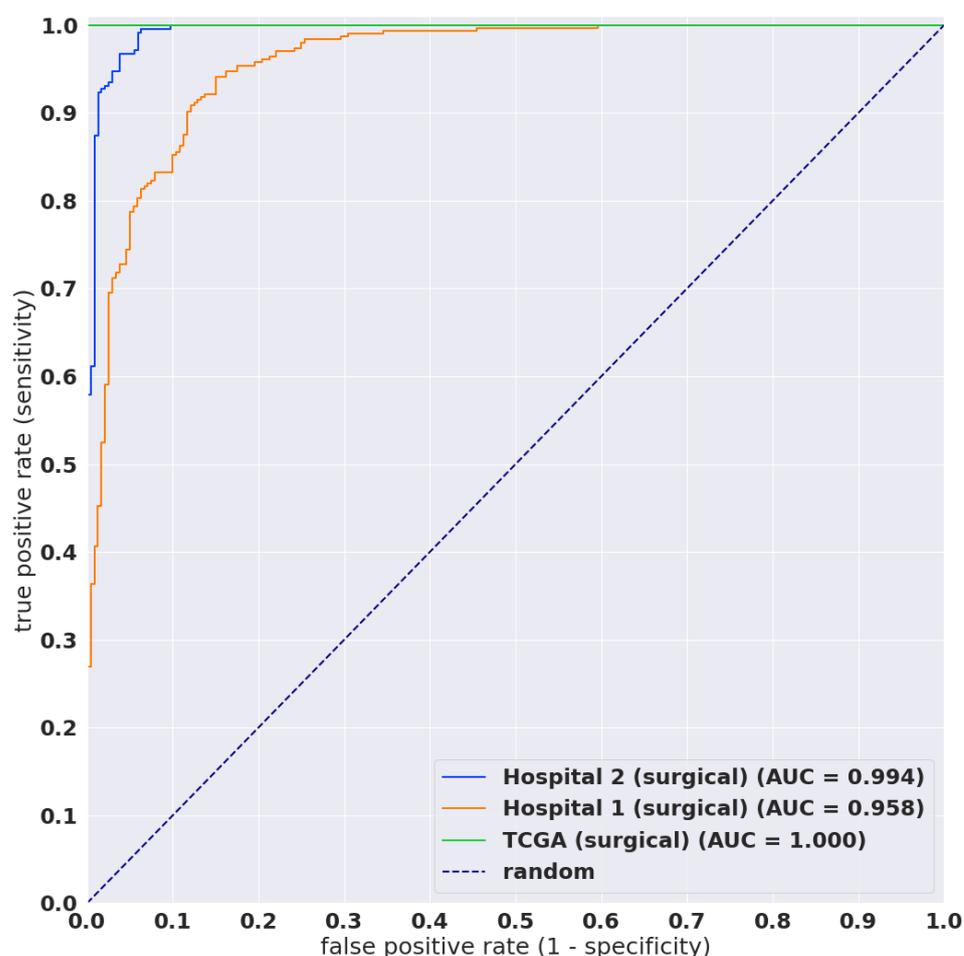


Figure 3. ROC curves on surgical test sets of the TL x10 512 B1 model.

Table 2. ROC and log loss results of the models on the biopsy and surgical test sets. The trained model names are referred to as TL <magnification> <tile size> <model size>.

Dataset	Model	ROC AUC	Log loss
Hospital 1 (biopsy)	Stomach ADC x10 512 [18]	0.853 [0.818, 0.884]	1.090 [0.955, 1.257]
	Colon ADC x10 512 [18]	0.691 [0.645, 0.735]	1.101 [0.966, 1.257]
	Lung carcinoma x10 224 [35]	0.664 [0.617, 0.710]	2.542 [2.184, 2.932]
	Pancreas ADC x10 224 [36]	0.800 [0.761, 0.835]	0.734 [0.661, 0.816]
	Stomach poorly-ADC x20 224 [37]	0.894 [0.867, 0.920]	0.548 [0.508, 0.587]
	Stomach signet ring x10 224 [34]	0.817 [0.790, 0.857]	0.895 [0.801, 0.976]
Hospital 1 (surgical)	TL x10 512 B1	0.980 [0.969, 0.991]	0.269 [0.201, 0.335]
	TL x10 224 B1	0.971 [0.957, 0.984]	0.258 [0.199, 0.317]
	TL x10 512 B3	0.979 [0.967, 0.989]	0.366 [0.284, 0.462]
	TL x20 512 B1	0.962 [0.945, 0.975]	0.285 [0.240, 0.346]
Hospital 2 (surgical)	TL x10 512 B1	0.994 [0.987, 0.998]	0.180 [0.139, 0.230]
	TL x10 224 B1	0.970 [0.956, 0.982]	0.399 [0.335, 0.476]
TCGA (surgical)	TL x10 512 B1	1.000 [1.000, 1.000]	0.274 [0.108, 0.332]
	TL x10 224 B1	0.997 [0.983, 1.000]	0.377 [0.245, 0.578]

Table 3. A breakdown of the subtypes of the false positives and true negatives in the biopsy test set using the TL model x10 using a classification threshold of 0.5.

	Subtype of Benign	Number of WSIs	%
False-positives (10 WSIs)	Fibroadenoma	10	100.0
True-negatives (223 WSIs)	Fibroadenoma	121	54.3
	Mastopathy	67	30.0
	Normal	24	10.8
	Fibrosis	6	2.7
	Ductal hyperplasia	2	0.9
	Granulation tissue	2	0.9
	Fat necrosis	1	0.4

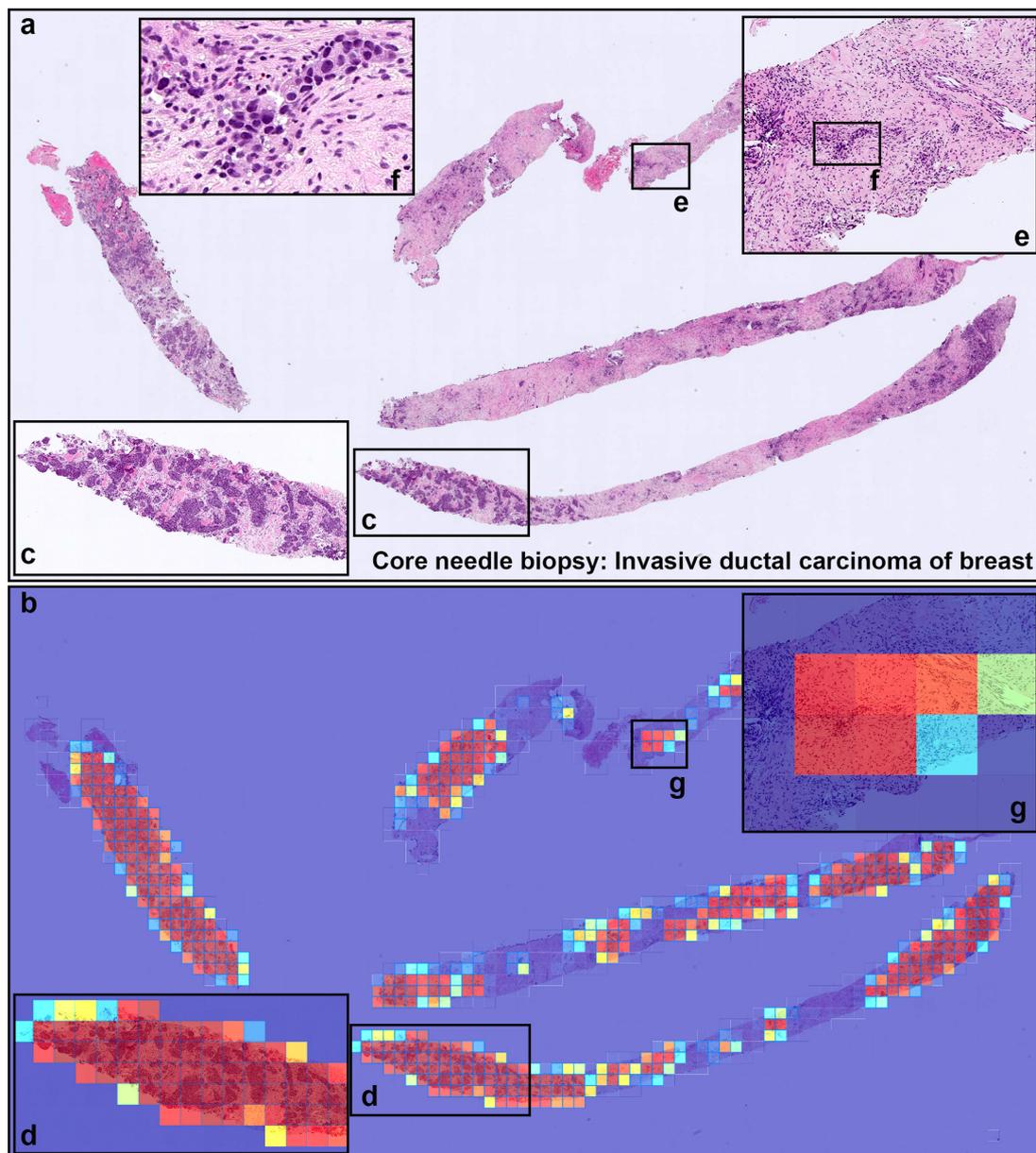


Figure 4. A representative true positive invasive ductal carcinoma (IDC) of breast from core needle biopsy test set. Heatmap images show true positive predictions of IDC cells (b) and they correspond, respectively, to H and E histopathology (a) using transfer learning from ImageNet model (magnification $\times 10$). Not only abundant IDC cells invading areas (c) but also a few IDC cells (e,f), heatmap images show appropriately true positive predictions (d,g).

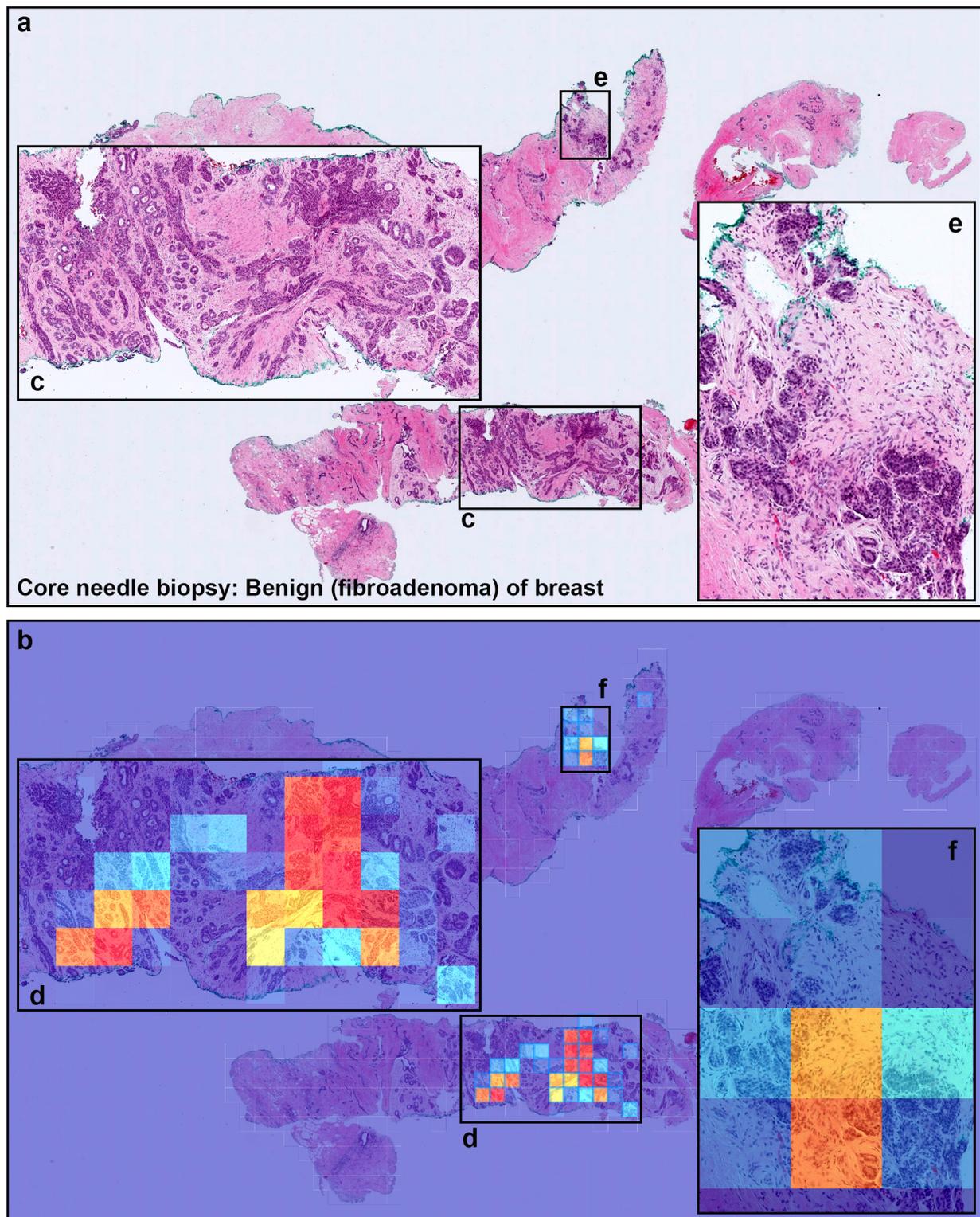


Figure 5. A representative example of invasive ductal carcinoma (IDC) false positive prediction output on a case from core needle biopsy test set. Histopathologically (a), this case is a benign lesion (fibroadenoma). Heatmap images (b,d,f) exhibited false positive prediction of IDC using transfer learning from ImageNet model (magnification $\times 10$). The ductular structures in fibroadenoma with a pericanalicular pattern (c–f) would be the primary cause of false positive due to its morphological analogous to ductular structures in IDC.

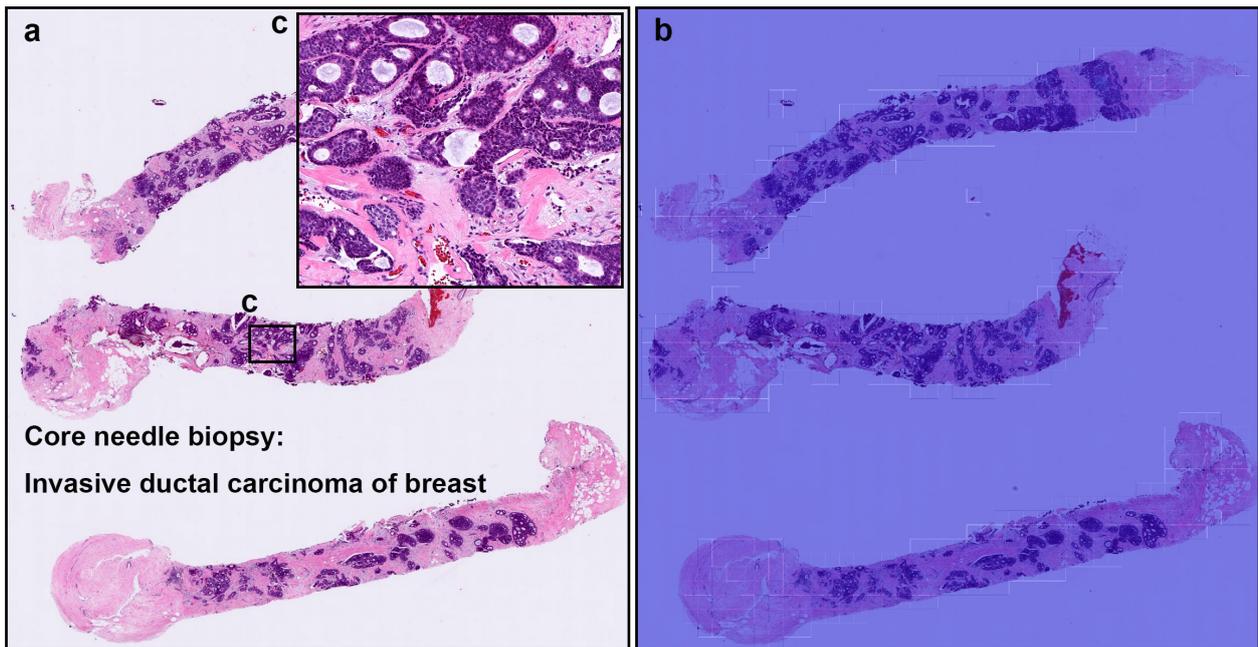


Figure 6. A representative false negative prediction output on a case from core needle biopsy test set. According to the histopathological report, this case (a,c) is an invasive ductal carcinoma (IDC). However, there are no true positive predictions of IDC cells on heatmap image (b).

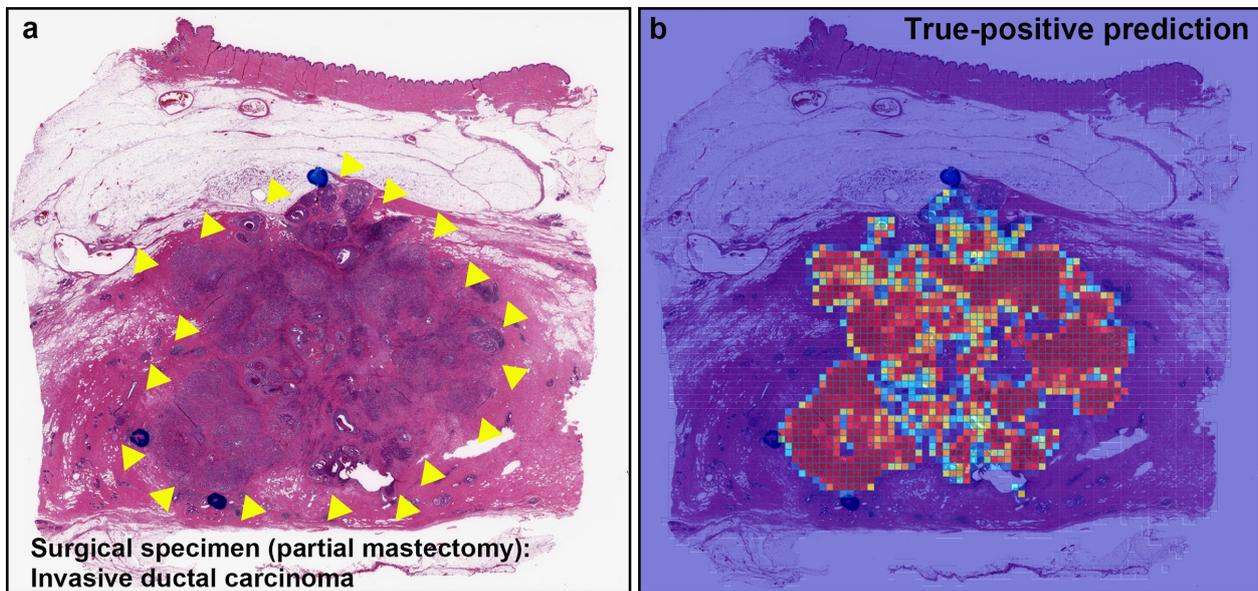


Figure 7. *Cont.*

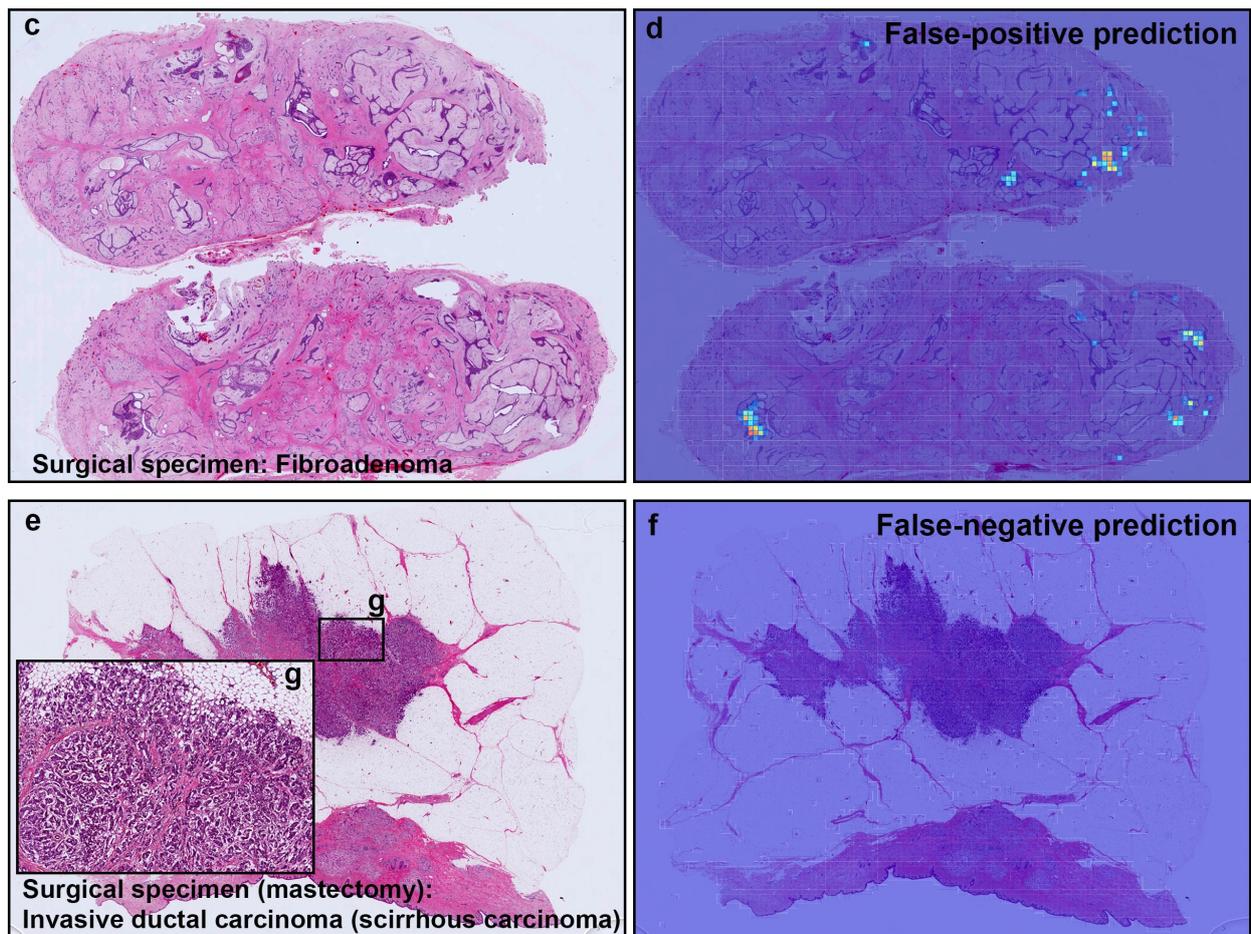


Figure 7. Representative true positive, false positive, and false negative prediction outputs on surgically resected specimens for invasive ductal carcinomas (IDCs) and fibroadenoma. Histopathologically, (a) has IDC; (c) is fibroadenoma; and (e) has IDC (scirrhous type); (b) shows true positive probability heatmap using transfer learning from ImageNet model (magnification $\times 10$) for IDC invading area which was corresponded to surgical pathologists marked area with blue-ink-dots (and yellow-triangles) (a); (d) exhibited false positive prediction of IDC in fibroadenoma. There are no true positive predictions of IDC cells on heatmap image (f) in scirrhous carcinoma of IDC (e).

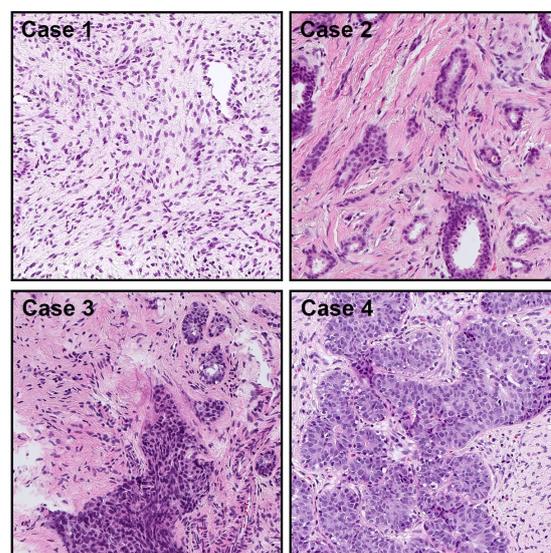


Figure 8. Cont.

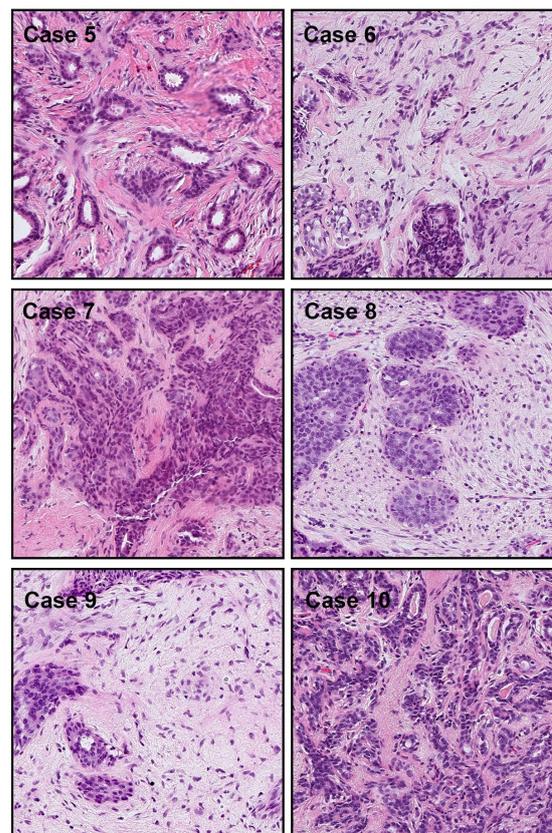


Figure 8. Representative tissue areas (Cases 1–10), without heatmap overlay, that were falsely predicted as IDC. There were 10 cases of false positive prediction outputs from the core needle biopsy test set. The false positive predictions are most likely due to the enlarged spindle shaped stromal cell nuclei with pleomorphism and tubules composed of cuboidal or low columnar cells with round uniform nuclei resting on a myoepithelial cell layer. This is morphologically analogous to invading single cells, ductular structures, and cancer stroma in IDC.

4. Discussion

In this study, we trained deep learning models for the classification of breast IDC in surgical and biopsy WSIs. We used weakly-supervised and transfer learning. We used the partial fine-tuning approach, which is fast to train. The best model achieved AUCs in the range of 0.96–0.98.

Overall, the EfficientNetB1 model trained at magnification $\times 10$ achieved slightly better results than $\times 20$ on the biopsy test set. In addition, using a larger tile size of 512×512 px achieved slightly better results than 224×244 px. Despite IDC morphology having some similarities with adenocarcinoma (ADC) [39], the application of models that classify ADC on other organs did not fully generalise to IDC. They have achieved lower AUC performances in the range 0.66 to 0.89. The stomach ADC had the highest AUC 0.85–0.89 when applied to breast IDC WSIs. While the results on the TCGA test set are high, it does not provide a proper evaluation in terms of potential false positives as there were only four benign cases.

All of the false positive cases in the biopsy test set were fibroadenomas (see Table 3). Fibroadenomas exhibit a wide range of morphology [38], and it could be that the variety was not fully represented in the training set, which only had 91 cases of fibroadenomas compared to the 131 in the test set. The false positive predictions with fibroadenomas are most likely due to the enlarged spindle shaped stromal cell nuclei with pleomorphism and tubules composed of cuboidal or low columnar cells with round uniform nuclei resting on a myoepithelial cell layer. This is morphologically analogous to invading single cells, ductular structures, and cancer stroma in IDC (see Figure 8).

One source of difficulty in creating a balanced set of the fibroadenomas varieties is that the diagnostic reports did not include a detailed description of fibroadenoma histology, making a simple random partition the only option. In addition, the test set had a larger proportion of fibroadenomas compared to other benign subtypes. Therefore, in future work, it would be important to investigate the histopathological typing of fibroadenomas in order to develop better deep learning models.

While in this study we relied on performing classification from WSI histopathology, in some cases, pathologists make use of immunohistochemistry markers to further confirm a diagnosis and guide treatment decisions. In particular, markers that allow for myoepithelial differentiation are useful for distinguishing between IDC and benign proliferations such as fibroadenoma [40]. This is because IDCs lack the myoepithelial cells that normally surround benign breast glands.

According to the guideline by General Rule Committee of the Japanese Breast Cancer Society [41], the pathological diagnosis of IDC is sufficient for core needle biopsy. Therefore, the application of a deep learning model, once properly validated, in a clinical setting would help pathologists in their diagnostic workflows potentially serving as a second reader during the screening process. It could also be used to sort cases in order of priority for review by the pathologists. On the other hand, surgical specimens tend to require further subtyping of IDC, so future work could look into developing models specifically for IDC subtype classification for surgical specimens.

5. Conclusions

In this study, we have trained deep learning models at two magnifications, $\times 10$ and $\times 20$, using transfer learning and weakly supervised learning for the classification of breast IDC in WSIs. We evaluated the models on four test sets (one biopsy and three surgical) achieving AUCs in the range 0.95 to 0.99. We have also compared the trained models to existing pre-trained models on different organs for adenocarcinoma classification and they have achieved lower AUC performances in the range 0.66 to 0.89 despite adenocarcinoma exhibiting some structural similarity to IDC. Therefore, performing fine-tuning on the breast IDC training set was beneficial for improving performance.

Author Contributions: F.K. and M.T. contributed equally to this work; F.K. and M.T. designed the studies, performed experiments, analysed the data, and wrote the manuscript; M.T. supervised the project. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The experimental protocol was approved by the ethical board of the International University of Health and Welfare (No. 19-Im-007) and Kamachi Group Hospitals (Shinkomonji and Shinkuki hospitals). All research activities complied with all relevant ethical regulations and were performed in accordance with relevant guidelines and regulations in all of the hospitals mentioned above.

Informed Consent Statement: Informed consent to use histopathological samples and pathological diagnostic reports for research purposes had previously been obtained from all patients prior to the surgical procedures at all hospitals, and the opportunity for refusal to participate in research had been guaranteed by an opt-out manner.

Data Availability Statement: Due to specific institutional requirements governing privacy protection, datasets used in this study are not publicly available. The TCGA data were obtained from TCGA-BRCA project and is available from <https://www.cancer.gov/tcga>, accessed on 6 June 2019.

Acknowledgments: We are grateful for the support provided by Takayuki Shiomi and Ichiro Mori at Department of Pathology, Faculty of Medicine, International University of Health and Welfare; Ryosuke Matsuoka at Diagnostic Pathology Center, International University of Health and Welfare, Mita Hospital; pathologists at Kamachi Group Hospitals (Fukuoka). We thank the pathologists who have been engaged in the annotation, reviewing cases, and pathological discussion for this study.

Conflicts of Interest: F.K. and M.T. are employees of Medmain Inc.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; 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 Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Kettritz, U.; Rotter, K.; Schreer, I.; Muraier, M.; Schulz-Wendtland, R.; Peter, D.; Heywang-Köbrunner, S.H. Stereotactic vacuum-assisted breast biopsy in 2874 patients: A multicenter study. *Cancer* **2004**, *100*, 245–251. [[CrossRef](#)]
3. Litherland, J.C. Should fine needle aspiration cytology in breast assessment be abandoned? *Clin. Radiol.* **2002**, *57*, 81–84. [[CrossRef](#)] [[PubMed](#)]
4. Collins, L.C.; Connolly, J.L.; Page, D.L.; Goulart, R.A.; Pisano, E.D.; Fajardo, L.L.; Berg, W.A.; Caudry, D.J.; McNeil, B.J.; Schnitt, S.J. Diagnostic agreement in the evaluation of image-guided breast core needle biopsies: Results from a randomized clinical trial. *Am. J. Surg. Pathol.* **2004**, *28*, 126–131. [[CrossRef](#)]
5. Yu, K.H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Ré, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. [[CrossRef](#)]
6. Hou, L.; Samaras, D.; Kurc, T.M.; Gao, Y.; Davis, J.E.; Saltz, J.H. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2424–2433.
7. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **2016**, *33*, 170–175. [[CrossRef](#)]
8. Litjens, G.; Sánchez, C.I.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Hulsbergen-Van De Kaa, C.; Bult, P.; Van Ginneken, B.; Van Der Laak, J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **2016**, *6*, 26286. [[CrossRef](#)]
9. Kraus, O.Z.; Ba, J.L.; Frey, B.J. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* **2016**, *32*, i52–i59. [[CrossRef](#)]
10. Korbar, B.; Olofson, A.M.; Miraflor, A.P.; Nicka, C.M.; Suriawinata, M.A.; Torresani, L.; Suriawinata, A.A.; Hassanpour, S. Deep learning for classification of colorectal polyps on whole-slide images. *J. Pathol. Inform.* **2017**, *8*, 30. [[PubMed](#)]
11. Luo, X.; Zang, X.; Yang, L.; Huang, J.; Liang, F.; Rodriguez-Canales, J.; Wistuba, I.I.; Gazdar, A.; Xie, Y.; Xiao, G. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J. Thorac. Oncol.* **2017**, *12*, 501–509. [[CrossRef](#)] [[PubMed](#)]
12. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [[CrossRef](#)] [[PubMed](#)]
13. Wei, J.W.; Tafe, L.J.; Linnik, Y.A.; Vaickus, L.J.; Tomita, N.; Hassanpour, S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci. Rep.* **2019**, *9*, 1–8.
14. Gertych, A.; Swiderska-Chadaj, Z.; Ma, Z.; Ing, N.; Markiewicz, T.; Cierniak, S.; Salemi, H.; Guzman, S.; Walts, A.E.; Knudsen, B.S. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci. Rep.* **2019**, *9*, 1483. [[CrossRef](#)] [[PubMed](#)]
15. Bejnordi, B.E.; Veta, M.; Van Diest, P.J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; Van Der Laak, J.A.; Hermsen, M.; Manson, Q.F.; Balkenhol, M.; et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **2017**, *318*, 2199–2210. [[CrossRef](#)] [[PubMed](#)]
16. Saltz, J.; Gupta, R.; Hou, L.; Kurc, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K.R.; Zhao, T.; Batiste, R.; et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **2018**, *23*, 181–193.
17. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Miraflor, A.; Silva, V.W.K.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **2019**, *25*, 1301–1309. [[CrossRef](#)]
18. Iizuka, O.; Kanavati, F.; Kato, K.; Rambeau, M.; Arihiro, K.; Tsuneki, M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci. Rep.* **2020**, *10*, 1–11. [[CrossRef](#)]
19. Bayramoglu, N.; Kannala, J.; Heikkilä, J. Deep learning for magnification independent breast cancer histopathology image classification. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2440–2445.
20. Sharma, S.; Mehra, R. Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—A Comparative Insight. *J. Digit. Imaging* **2020**, *33*, 632–654. [[CrossRef](#)]
21. Hameed, Z.; Zahia, S.; Garcia-Zapirain, B.; Aguirre, J.J.; Vanegas, A.M. Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models. *Sensors* **2020**, *20*, 4373. [[CrossRef](#)]
22. Mi, W.; Li, J.; Guo, Y.; Ren, X.; Liang, Z.; Zhang, T.; Zou, H. Deep Learning-Based Multi-Class Classification of Breast Digital Pathology Images. *Cancer Manag. Res.* **2021**, *13*, 4605–4617. [[CrossRef](#)]

23. Sohail, A.; Khan, A.; Nisar, H.; Tabassum, S.; Zameer, A. Mitotic nuclei analysis in breast cancer histopathology images using deep ensemble classifier. *Med Image Anal.* **2021**, *72*, 102121. [[CrossRef](#)] [[PubMed](#)]
24. Wetstein, S.C.; Stathonikos, N.; Pluim, J.P.W.; Heng, Y.J.; ter Hoeve, N.D.; Vreuls, C.P.H.; van Diest, P.J.; Veta, M. Deep learning-based grading of ductal carcinoma in situ in breast histopathology images. *Lab. Investig.* **2021**, *101*, 525–533. [[CrossRef](#)]
25. Kanavati, F.; Tsuneki, M. Partial transfusion: On the expressive influence of trainable batch norm parameters for transfer learning. *arXiv* **2021**, arXiv:2102.05543.
26. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning PMLR, Long Beach, CA, USA, 24 May 2019; pp. 6105–6114.
27. Kanavati, F.; Toyokawa, G.; Momosaki, S.; Rambeau, M.; Kozuma, Y.; Shoji, F.; Yamazaki, K.; Takeo, S.; Iizuka, O.; Tsuneki, M. Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci. Rep.* **2020**, *10*, 1–11. [[CrossRef](#)]
28. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
29. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
30. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: [tensorflow.org](https://www.tensorflow.org) (accessed on 24 January 2021).
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
32. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
33. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
34. Kanavati, F.; Tsuneki, M. A deep learning model for gastric diffuse-type adenocarcinoma classification in whole slide images. *arXiv* **2021**, arXiv:2104.12478.
35. Kanavati, F.; Toyokawa, G.; Momosaki, S.; Takeoka, H.; Okamoto, M.; Yamazaki, K.; Takeo, S.; Iizuka, O.; Tsuneki, M. A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. *Sci. Rep.* **2021**, *11*, 1–14.
36. Naito, Y.; Tsuneki, M.; Fukushima, N.; Koga, Y.; Higashi, M.; Notohara, K.; Aishima, S.; Ohike, N.; Tajiri, T.; Yamaguchi, H.; et al. A deep learning model to detect pancreatic ductal adenocarcinoma on endoscopic ultrasound-guided fine-needle biopsy. *Sci. Rep.* **2021**, *11*, 1–8. [[CrossRef](#)]
37. Kanavati, F.; Ichihara, S.; Rambeau, M.; Iizuka, O.; Arihiro, K.; Tsuneki, M. Deep learning models for gastric signet ring cell carcinoma classification in whole slide images. *Technol. Cancer Res. Treat.* **2021**, *20*, 15330338211027901. [[CrossRef](#)] [[PubMed](#)]
38. Kuijper, A.; Mommers, E.C.; van der Wall, E.; van Diest, P.J. Histopathology of fibroadenoma of the breast. *Am. J. Clin. Pathol.* **2001**, *115*, 736–742. [[CrossRef](#)] [[PubMed](#)]
39. Makki, J. Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clin. Med. Insights Pathol.* **2015**, *8*, CPath-S31563. [[CrossRef](#)] [[PubMed](#)]
40. Zaha, D.C. Significance of immunohistochemistry in breast cancer. *World J. Clin. Oncol.* **2014**, *5*, 382. [[CrossRef](#)]
41. Tsuda, H. Histological classification of breast tumors in the General Rules for Clinical and Pathological Recording of Breast Cancer. *Breast Cancer* **2020**, *27*, 309–321. [[CrossRef](#)]