Outcome and biomarker supervised deep learning for survival prediction in two multicenter breast cancer series

Dmitrii Bychkov a,b,⁎, Heikki Joensuu b,c, Stig Nordin d, Aleksei Tiulpin e,f,g, Hakan Küçükel a,b, Mikael Lundin a, Harri Sihto d, Jorma Isola h, Tiina Lehtimäki i, Pirkko-Liisa Kellokumpu-Lehtinen j, Karl von Smitten k, Johan Lundin a,b,l, Nina Linder a,b,m

a Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
b ICAN Digital Precision Cancer Medicine Program, Finland
c Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
d Department of Oncology, Tampere University Hospital, Tampere, Finland
e Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
f Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland
g Ailean Technologies Oy, Oulu, Finland
h Department of Cancer Biology, BioMediTech, University of Tampere, Tampere, Finland
i Eira Hospital, Helsinki, Finland
j Helsinki University Hospital, Helsinki, Finland
k Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden
l Department of Women’s and Children’s Health, International Maternal and Child Health, Uppsala University, Uppsala, Sweden

A R T I C L E   I N F O

Keywords:
Breast cancer
Convolutional neural networks
Digital pathology
ERBB2 gene
Estrogen receptor
Multitask deep learning
Outcome prediction

A B S T R A C T

Background: Prediction of clinical outcomes for individual cancer patients is an important step in the disease diagnosis and subsequently guides the treatment and patient counseling. In this work, we develop and evaluate a joint outcome and biomarker supervised (estrogen receptor expression and ERBB2 expression and gene amplification) multitask deep learning model for prediction of outcome in breast cancer patients in two nationwide multicenter studies in Finland (the FinProg and FinHer studies). Our approach combines deep learning with expert knowledge to provide more accurate, robust, and integrated prediction of breast cancer outcomes.

Materials and methods: Using deep learning, we trained convolutional neural networks (CNNs) with digitized tissue microarray (TMA) samples of primary hematxylin-eosin-stained breast cancer specimens from 693 patients in the FinProg series as input and breast cancer-specific survival as the endpoint. The trained algorithms were tested on 354 TMA patient samples in the same series. An independent set of whole-slide (WS) tumor samples from 674 patients in another multicenter study (FinHer) was used to validate and verify the generalization of the outcome prediction across both studies. The algorithms learned 28 visual features from breast cancer tissue samples assessed by pathologists.

Results: The multitask algorithm achieved a hazard ratio (HR) of 2.0 (95% confidence interval [CI] 1.30–3.00), P < 0.001, c-index of 0.59 on the test set of FinProg patients, and an HR of 1.7 (95% CI 1.2–2.6), P = 0.003, c-index of 0.57 on the WS tumor samples from 674 patients in the independent FinHer series. The multitask CNN remained the statistically independent predictor of survival in both test sets when adjusted for histological grade, tumor size, and number of mitoses, tubules, nuclear pleomorphism, tumor-infiltrating lymphocytes, and necrosis was performed on TMA samples in the FinProg test set by a pathologist and combined with deep learning-based outcome prediction in a multitask algorithm.

Conclusions: A multitask deep learning algorithm supervised by both patient outcome and biomarker status learned features in basic tissue morphology predictive of survival in a nationwide, multicenter series of patients with breast cancer. The algorithms generalized to other independent multicenter patient series and whole-slide breast cancer samples and provide prognostic information complementary to that of a comprehensive series of established prognostic factors.

⁎ Corresponding author at: Institute for Molecular Medicine Finland FIMM, Nordic EMBL Partnership for Molecular Medicine, P. O. Box: 20, FI-00014 University of Helsinki, Helsinki, Finland.
E-mail address: dmitrii.bychkov@helsinki.fi (D. Bychkov).

http://dx.doi.org/10.4103/jpi.jpi_29_21
Received 22 April 2021; Received in revised form 10 June 2021; Accepted 20 June 2021
Available online 20 December 2022
© 2022 Published by Elsevier Inc. on behalf of Association for Pathology Informatics. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

In this study, we suggest a novel approach for extraction of cancer outcome-related information from tissue morphology by joint outcome and biomarker supervised deep learning with convolutional neural networks (CNNs). This technique is known as multitask learning and has not been previously applied to outcome prediction in breast cancer using conventional hematoxylin-eosin (H and E) stained digitized tissue specimens. It has been demonstrated that a multitask approach can improve the accuracy of classification of breast tissue samples according to histologic type and grade of differentiation as well as diagnosis of breast cancer in mammograms.

In breast cancer, the expression of estrogen receptors (ER), as well as the expression and gene amplification of ERBB2 (erb-b2 receptor tyrosine kinase 2, also known as HER2) guide the selection of treatment. Previous studies demonstrate that the ER and ERBB2 status can be predicted directly from HE stained breast cancer tissue samples and that the tissue morphological features predictive of the ERBB2 status also predict patient outcome. In addition, it has been shown that CNNs can be trained to predict survival in breast cancer directly from the tissue morphology, supervised by patient outcome. Therefore, we hypothesized that a combination of biomarker and outcome-supervised training with a multitask approach could improve the accuracy of outcome prediction in breast cancer.

To assess if outcome and biomarker supervised multitask CNNs can learn tissue-based features that are independent of established prognostic factors, a series of tissue characteristics including histological grade (with subfactors mitotic figures, nuclear pleomorphism, tubule formation), tumor necrosis, axillary lymph node status and tumor size were included in multivariate prognostic models. We evaluated how these characteristics, which were determined visually by a human expert, and the prognostic information extracted by CNN-based multitask learning could complement each other in breast cancer outcome prediction. In the current study, we trained the algorithms with images from tissue microarray (TMA) samples from a nationwide patient series and then validated the results on whole slide tissue specimens from another multicenter trial. Our aim was to validate the generalization of the deep learning algorithms for outcome prediction when applied to breast cancer from an independent patient series.

2. Materials and methods

2.1. Patient series

The study was based on cancer tissue samples, clinicopathological data, and follow-up data from two independent breast cancer series: The FinProg series (which consists of the original FinProg series and the FinProg validation series) and the FinHer clinical trial series (ISRCTN76560285). The original FinProg patient series with data from 2936 patients, is a nationwide series that includes approximately 50% of all women diagnosed with breast cancer in Finland in 1991 or 1992 and covers most (93%) of the patients with breast cancer diagnosed within five selected geographical regions [Supplementary Fig. 1]. The FinProg validation series consists of 565 patients diagnosed mainly in the Helsinki metropolitan region who were treated at the Departments of Surgery and Oncology, Helsinki University Hospital, from 1987 to 1990. The outcome and cause of death data (breast cancer-specific survival [BCSS]) were retrieved from the Finnish Cancer Registry and Statistics Finland. Corresponding clinical information and pathologic tumor characteristics, including cancer histological grade, tumor size in centimeters, and axillary lymph node status were available from the hospital records. Tumor TMAs were prepared from each patient's representative formalin-fixed paraffin-embedded breast cancer samples. Amplification of the ERBB2 gene was quantified by chromogenic in situ hybridization (CISH) on TMA core sections as described previously, and ER expression was determined by immunohistochemistry. A total of 1047 FinProg patients with one TMA image per patient were split into a training and tuning set (n = 693) and an internal test set (n = 354) [Table 1 and Supplementary Fig. 1]. The median follow-up time of patients included in the training and tuning set was 15.5 years.

The FinHer trial (ISRCTN76560285) was an open-label multicenter randomized trial that included 1010 patients in Finland in 2000-2003. Eligible women were ≤ 65 years of age, had undergone breast surgery with axillary nodal dissection, and had either axillary lymph node-positive or high-risk node-negative cancer [Supplementary Fig. 2]. Breast cancer ER and ERBB2 expression were determined by immunohistochemistry according to institutional guidelines. For patient samples considered positive for ERBB2 expression by immunohistochemistry (either 2+ or 3+ on a scale from 0 to 3+), ERBB2 gene amplification status was determined by CISH. Breast cancers with > 6 gene copies were considered ERBB2-positive. Patients were randomly assigned to receive three cycles of doxetaxel or vinorelbine, followed in both groups by three cycles of fluorouracil, epirubicin, and cyclophosphamide. The 232 (23.0%) patients with ERBB2-positive cancer underwent a second randomization either to receive concomitant intravenous trastuzumab for 9 weeks or to not receive trastuzumab. One patient with overt distant metastases at the time of random assignment was excluded from survival analyses. The primary end-point for the FinHer trial participants was distant disease-free survival (DDFS), defined as the time from randomization to the detection of distant metastasis. The median follow-up time was 5.2 years after random assignment. A total of 712 HE-stained whole-slide images (WSIs), one slide per FinHer patient were used as an external test set, not used for training or tuning the algorithms [Table 1 and Supplementary Fig. 2].

2.2. Ethics approval

The use of the FinProg patient series and the clinical data was approved by the operative Ethics Committee of the Hospital District of Helsinki and Uusimaa (94/13/03/02/2012), and the National Supervisory Authority for Welfare and Health (Valvira) approved the use of human tissues (7717/06.01.03/01/2015). Profiling of tumors from the FinHer patient series was approved by the institutional review board of the Helsinki University Hospital (HUS 177/13/03/02/2011).

2.3. Annotation of tissue images

Mitotic figures, nuclear pleomorphism, and tubule formation were assessed by a pathologist (S.N.) on 354 TMA spot images from the FinProg test set. These expert-derived features were further combined into a TMA-based histological grade according to a modification of the established breast cancer grading system [Supplementary Table 1]. Scores 3–5, 6–7, and 8–9 formed grades I, II, and III, respectively. Tissue necrosis and tumor-infiltrating lymphocytes (TILs) were also assessed on the same set of FinProg TMA images. Further, a visual risk score (VRS) was determined by a pathologist (S.N.), such that the patients were assigned to a low-risk or a high-risk group, based on the morphology of the corresponding TMA samples.

2.4. Image preprocessing and augmentation

Images of TMA samples from the FinProg series (average size 3500 × 3500 pixels) were available in a Portable Network Graphics format extracted from WSIs scanned with a whole slide scanner (Pannoramic 250 FLASH, 3DHISTECH Ltd., Budapest, Hungary) and the FinHer samples as original whole-slide image files (MRXS) digitized with the same scanner [Supplementary Material]. Tiles of 950 × 950 pixels [209 × 209 µm with 0.22 µm pixel size, Supplementary Material] were extracted from the FinHer WSIs and saved in a JPEG format. Both the FinProg TMA images and FinHer image tiles were color-normalized to adjust for HE staining variation across the tissue samples.

During training on FinProg images, we extracted square crops from a random location in the TMA spot images. One crop of size 950 × 950 pixels per TMA spot was extracted at each epoch. Thus, at every epoch, the
networks were supplied with a different set of crops that originated from various locations of the TMA spots included in the training set. On the fly, data augmentation was applied to the FinProg TMA images during training. Image up/down-scale (0%–30%), rotation (±90°), shear (0%–20%), and gamma correction (0%–30%) were randomly applied to the TMA crops [Supplementary Material].

2.5. Network architecture and training

We built the deep learning model around a ResNet21 CNN backbone. The backbone constitutes a stack of convolutional layers and outputs three-dimensional arrays, i.e., feature maps. These feature maps are globally average pooled to produce a feature vector of a fixed size. Thereby, global average pooling allows to input images of arbitrary size into the outcome prediction pipeline. Finally, the feature vector is passed through a fully connected layer to predict a corresponding continuous-value risk score, associated with the input image. The GuanRank,22 an on-parametric ranking-based technique was used to transform time-to-event data into a linear space of hazard ranks representing BCSS for each patient. Thereby, the outcome prediction was turned into a regression task with the mean absolute error as loss function. Mean squared error loss was used to evaluate the generalization of the models trained on the FinProg TMA sample images we employed two independent test sets: The FinProg test-set patients that we refer to as the internal test set and the FinHer patient series that we did not use for training at all. In both sets, we averaged outputs from the five models trained in cross-validation to reduce the variance of the CNNs and boost the prediction accuracy.

2.6. Inference procedure

To evaluate the generalization of the models trained on the FinProg TMA sample images we employed two independent test sets: The FinProg test-set patients that we refer to as the internal test set and the FinHer patient series that we did not use for training at all. In both sets, we averaged outputs from the five models trained in cross-validation to reduce the variance of the CNNs and boost the prediction accuracy.

2.7. Statistical analysis

The concordance between the predicted risk score (CNN output) and the actual time-to-event data (follow-up time and censor status) was estimated with the concordance index (c-index) in the patients included in the test sets. We applied Cox Proportional Hazards (PH) univariate survival regression to derive hazard ratios (HR; effect size) associated with the risk score predicted by the CNNs and other clinicopathological variables. In addition, Cox PH multivariate regression was performed to check the independence of the variables in prediction of the risk score. The log-rank test was used to compare survival distributions between two patient subgroups.

3. Results

3.1. Multitask learning and outcome prediction accuracy in the FinProg series

We trained CNNs to extract prognostic information from the breast cancer TMA samples in the FinProg series [Fig. 1]. We used TMA images from 693 FinProg patients to train the algorithm in a five-fold cross-validation and then applied the trained models to a test set of 354 FinProg patients. The “Solo” models that were supervised with outcome data only (i.e., the GuanRank value) achieved an HR of 1.7 (95% confidence interval [CI] 1.10–2.60) in a univariate Cox PH regression, \( P = 0.009 \) and concordance index (c-index) of 0.57 [Table 2]. Models trained in a multitask fashion, i.e., predicting ER and ERBB2 status together with outcome achieved an HR of 2.0 (95% CI 1.30–3.00), \( P < 0.001 \), and an accuracy as measured by the c-index of 0.59 [Table 2]. Examples of high-risk and low-risk FinProg patient TMA samples are presented in Supplementary Fig. 3.
3.2. Morphological characteristics of tumors assessed on tissue microarray samples predict patient survival

We examined whether the subcomponents of histological grade, i.e., mitotic figures, nuclear pleomorphism, and tubule formation predict survival of patients in the FinProg series when assessed by a pathologist viewing the TMA images. Univariate Cox PH regression showed that all three features were predictive of BCSS. Marked nuclear pleomorphism had an HR of 3.00 (95% CI 1.34–6.70), \( P = 0.008 \), c-index of 0.59; low tubules formation had an HR of 2.20 (95% CI 1.10–4.60); high mitotic count reached an HR of 2.00 (95% CI 1.10–3.60) (Table 2). The TMA-based grading had an HR of 3.00 (95% CI 1.50–6.10), \( P = 0.002 \), and a c-index of 0.60 on the FinProg test set (Table 2). The original histological grading assessed on whole-slides (WS grade) by pathologists at the time of diagnosis demonstrated an HR of 4.00 (95% CI 2.00–8.30), \( P < 0.001 \), and a c-index of 0.64. The presence of necrotic tissue was associated with an HR of 5.00 (95% CI 2.40–10.00), \( P < 0.001 \), whereas a higher number of TILs was not a statistically significant predictor of survival in a univariate Cox PH regression [Table 2]. The VRS reached an HR of 1.80 (95% CI 1.20–2.70), \( P = 0.004 \), and a c-index of 0.58.

3.3. Deep learning combined with expert visual assessment of tissue samples

To evaluate how the deep learning-based outcome prediction can complement visual tissue assessment, we first combined “solo” and multitask CNN models with visual TMA-based histological grading. The multitariate (TMA grade + CNN) Cox PH regression showed that the multitask CNN was an independent predictor of BCSS when adjusted for the visual TMA-based histological grade with an HR of 1.7 (95% CI 1.10–2.70), a \( P = 0.017 \), and a c-index of 0.63. A similar c-index (0.63) was observed when the multitask CNN was combined with the VRS. Importantly, the “solo” CNN was not a statistically significant predictor of BCSS when adjusted for the TMA-based histological grade and for the VRS.

3.4. Generalization to independent series whole slide samples

To evaluate generalization of the proposed approach, the CNNs trained on the TMA samples from the FinProg patient series were applied to WSIs
from the independent FinHer patient series. Univariate Cox PH regression showed that both multitask, and “solo” CNN models were statistically significant predictors of DDfFS in patients from the FinHer series (n = 674). The “solo” model reached an HR of 1.8 (95% CI 1.2–2.7), a P = 0.002 and a c-index of 0.57. The multitask model achieved an HR of 1.7 (95% CI 1.2–2.6), P = 0.003 and a c-index 0.57. We then evaluated both of the models in a multivariate Cox PH regression adjusted for the WS histological grade and observed that both of the models were statistically significant predictors of survival, independent of histological grade on WSs [Table 3]. The “solo” model reached an HR of 1.7 (95% CI 1.1–2.5), a P = 0.009 and a c-index of 0.60 in a multivariate Cox PH analysis, whereas the multitask CNN reached an HR of 1.5 (95% CI 1.0–2.3), a P = 0.033 and a c-index of 0.59 [Table 3].

4. Conclusions

Our study demonstrates the feasibility of breast cancer outcome prediction using a multitask deep learning approach across two multicenter patient series. We show that the algorithms trained on one patient series (FinProg) can generalize to an independent patient series (FinHer). Although several studies have shown that outcome supervised deep learning can extract significant prognostic information from tumor morphology in breast cancer, they are constrained to the analysis of single-center series. To our knowledge, this work is the first to explore generalization of the method when applied to whole-slide breast cancer tissue images from an independent multicenter patient series.

With images of H and E-stained tumor tissue samples as the input, we applied both outcome and biomarker supervised learning to extract predictive information encoded in the tumor morphology. Our best multitask algorithm achieved an HR of 2.0 and a c-index of 0.59 in predicting BCSS in the FinProg test set patients. Moreover, we demonstrated that the multitask approach allows extraction of image features that remain independent of the pathologist-derived features such as mitoses, nuclear pleomorphism, tubules, and necrosis. In contrast to the “solo” training, the multitask deep learning-based risk score was a significant predictor of breast cancer-specific survival after adjustment for tumor size and axillary lymph nodes status in the FinProg series. Interestingly, we observed that the information extracted through visual assessment of TMA images by a pathologist and by the CNNs together could ultimately increase prognostic accuracy to a c-index of 0.66. We recognize that the multitask approach did not demonstrate an increased accuracy in the external FinHer WS samples, as compared to the “solo” model. Potential reasons could be different endpoints used in the FinProg and the FinHer, a relatively short follow-up time, and significant heterogeneity introduced through analysis of whole slide tissue samples within the FinHer series. In the current study, we did not explore the tile size sampling effect on the performance of the models, since the training set comprised images of TMAs. Training with tiles smaller than 950 × 950 pixels (209 × 209 μm) would limit the contextual information further and likely lead to a reduced performance. On the other hand, larger tiles would lead to increased morphological heterogeneity and inclusion of non-tumor tissue areas if extended to WSIs, and likely require even larger datasets than the sample series available in the current study if training is done with sample-level labels i.e., weakly supervised learning. Systematic evaluation of sampling strategies has to be studied separately. Taken together, a deep learning model trained on TMA samples stained for basic morphology (HE) and supervised by outcome and biomarker status based complemented visual tissue assessment of established tissue entities by a pathologist in the prediction of patient outcome.

In one of the first studies27 to address breast cancer outcome prediction with machine learning applied to basic tissue morphology, the authors used regularized logistic regression and image features from breast cancer epithelium and stroma. This approach reached HRs of 1.54–1.78 in two patient populations as estimated by a multivariate Cox PH regression in prediction of overall survival. These effect sizes are roughly at the same level as the HRs of 1.5–2.0 that were measured in prediction of BCSS and DDfFS in the current patient series. Another study2 used a deep learning approach to predict BCSS in one of the series also used in the current study (the FinProg series). The machine learning-based predictor reached an HR of 2.04 in a test set of 431 patients, but the approach was not validated on independent data. In a study that addressed morphology-based cancer survival prediction in multiple cancer types,4 the authors trained a deep learning model on 488 WS breast cancer samples from the Cancer Genome Atlas (TCGA) project. An HR of 2.86 was achieved in a multivariate Cox PH analysis on 250 holdout patients from the same TCGA patient cohort without cross-validation. Results so far suggest that significant prognostic information can be extracted from basic tissue morphology by the use of machine learning, but that effect sizes do not yet exceed those for some of the established prognostic tissue features currently assessed visually by experts. It remains to be established if the prognostic accuracy can be further improved by training and validating algorithms based on WSIs that better represent tissue heterogeneity as compared to TMAs.

Limitations related to our study include that BCSS was used in training of the algorithm on the FinProg data whereas DDfFS was used as an endpoint for evaluation on the FinHer series. Although a strong correlation
has been shown between disease-free and overall survival in studies on early breast cancer, the strength of correlation between BCSS and DFS remains to be established. Additionally, the tissue samples used in our study were centrally scanned using the same instrument. Thus, possible image variations due to the scanning hardware were eliminated but the generalizability of the method to samples digitized with other similar whole slide scanners have to be addressed in future studies.

In future research, the prognostic accuracy and generalization of the deep learning models can be further improved by exposing deep learning algorithms to datasets that cover an even larger spectrum of variations of tissue morphologies, including training on WSIs. Quantification of conventional prognostic features using machine learning algorithms instead of visual assessment as in the current study could further improve accuracy, consistency, and reproducibility of outcome prediction. Previous studies have demonstrated a good performance of machine learning algorithms in counting mitosis, quantifying tumor-infiltrating immune cell, assessing the grade of tumor differentiation, and tissue necrosis. A combination of computationally quantified conventional prognostic features with features learned through end-to-end outcome supervised learning should be addressed in future studies.

Our findings indicate that outcome and biomarker supervised deep learning models for breast cancer outcome prediction generalize to patient samples from an independent multicenter series. Integrative techniques such as multitask deep learning can extract image features that remain statistically independent of established prognostic factors in breast cancer. Hence, established prognostic features and features learned through machine learning approaches can complement each other and lead to more accurate and interpretable tumor tissue analysis for patient cancer outcome prediction.

Financial support and sponsorship

The study was supported by the Sigrid Juselius Foundation, the Biomedicum Helsinki Foundation, the Orion-Pharmos Research Foundation, Finska Läkaresällskapet, Medicinska Understödsföreningen Liv och Hälsa, Stiftelsen Dorothea Olivia, Karl Walter och Jarl Walter Perklén minne, K. Albin Johanssons Stiftelse, ICAN Digital Precision Cancer Medicine Flagship, and HiLIFE Helsinki Institute of Life Sciences.

Availability of data and materials

The data that support the findings of this study were used under a license for the current study, and some restrictions apply to their availability. The data are available from the authors upon reasonable request and with permission from the University of Helsinki.

Declaration of Competing Interest

Johan Lundin and Mikael Lundin are the founders and co-owners of Aiforia Technologies Oy, Helsinki, Finland. Heikki Joensuu is employed by Orion Pharma, serves as the Chairman of the Advisory Board of Neuron Therapeutics, has received funds from Neuron Therapeutics and owns stock of Orion Pharma and Sartar Therapeutics. Aleksei Tiulpin is a co-founder, shareholder, and CTO of Aliean Technologies Oy. The other authors have no conflicts of interest.

Acknowledgments

We would like to thank the Digital Microscopy and Molecular Pathology unit at Institute for Molecular Medicine Finland FIMM, University of Helsinki, supported by the Helsinki Institute of Life Science and Biocenter Finland for providing slide scanning services.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.4103/jpi.jpi_29_21.

References

1. Mohadesary P, Yousefi S, Angmad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci U S A 2018;115: E2070.DOI:10.1073/pnas.1720979115.
2. Bychkov D, Linder N, Turuki R, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep 2018;8:3395.
3. Turuki R, Bychkov D, Lundin M, et al. Breast cancer outcome prediction with tumour tissue images and machine learning. Breast Cancer Res Treat 2019;177:41-52.
4. Walczyn E, Steiner DF, Xu Z, et al. Deep learning-based survival prediction for multiple cancer types using histopathology images. PLoS One 2020;15:e0233678.
5. Argyriou A, Egeniou T, Pontil M. Multi-task feature learning. Adv Neural Inf Process Syst 2007;19:41-48.
6. Xipeng P, Li L, Yang H, et al. Multi-task deep learning for fine-grained classification and grading in breast cancer histopathological images. Multimodal Tt时限 Appl 2020;79; https://doi.org/10.1063/1.5139269.
7. Samala RK, Chan HP, Hadijskimi LM, Helvie MA, Cha KH, Richter CD. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys Med Biol 2017;62:8984-8990.
8. Shamai G, Binenbaum Y, Slosberg R, Dues I, Gil Z, Kaminer R. Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer. JAMA Netw Open 2019;2, e197700.
9. Rawat RR, Ortega I, Roy P, et al. Deep Learned Tissue ‘Fingerprints’ Classify Breast Cancers by ER/PR/Her2 Status from H and E Images. Springer Science and Business Media (LLC). Sci Rep 2020;10, https://doi.org/10.1038/s41598-020-64156-4. (Last access date 2021 Jun 01).
10. Naik N, Madani A, Esteve A, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-line H and E stains. Nat Commun 2020;11:5727.
11. Bychkov D, Linder N, Tiulpin A, et al. Deep learning identifies morphological features in breast cancer predictive of ERBB2 status and trastuzumab treatment efficacy. Sci Rep 2021;11:4037.
12. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. 1. the value of histo-pathological grade in breast cancer: experience from a large study with long-term follow-up. Pathobiology 1991;19:403-410.
13. WHO Classificational Tumours Editorial Board. WHO Classification of Breast Tumours: WHO Classification of Tumours. Vol. 2.World Health Organization. 2019.
14. Joensuu H, Isola J, Lundin M, et al. Amplification of erbB2 and erbB2 expression are superior to estrogen receptor status as risk factors for distant recurrence in pT1N0M0 breast cancer: a nationwide population-based study. Clin Cancer Res 2009;15:923-930.
15. Lundin J, Lundin M, Isola J, Joensuu H. A web-based system for individualised survival estimation in breast cancer. BMI 2005:526-29.
16. Joensuu H, Kellokumpu-Lehtinen PL, Bono P, et al. Adjuvant doctaxel or vinorelbine with or without trastuzumab for breast cancer. N Engl J Med 2006;354:809-820.
17. Joensuu H, Tiira I, Kaaja H, et al. Risk for distant recurrence of breast cancer detected by mammography screening or other methods. JAMA 2004;292:1064-1073.
18. Joensuu H, Pettit B, Vesa K, et al. Fluorouracil, epirubicin, and cyclophosphamide with either doctaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments for breast cancer: final results of the Finher trial. J Clin Oncol 2009;27:5685-5692.
19. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359-377.
20. Maeneno M, Niemhammer M, Marron J, et al. A method for normalizing histology slides for quantitative analysis. Proceedings of the Sixth IEEE International Conference on Symposium on Biomodal Imaging: From Nano to Macro; 2009. p. 1107-1110.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv e-prints 2015; https://arxiv.org/abs/1512.03385. Available from: http://arxiv.org/abs/1512.03385 (Last access date 2021 Jun 01).
22. Huang Z, Zhang H, Box J, et al. Complete hazard ranking to analyze right-censored data: an ALS survival study. PLoS Comput Biol 2017;13, e1005887.
23. Paszke A, Gross S, Mass F, et al. PyTorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R, eds. Advances in Neural Information Processing Systems 32. Curran Associates; Inc. 2019. p. 8024-8035.
24. Kingma DP, Ba J, adam: a method for stochastic optimization. arXiv e-prints 2017. p. arXiv:1412.6980.
25. Lin TY, Goyal P, Girshick RB, He K, Dollár P. Focal loss for dense object detection. 2017.p. arXiv: 1512.03385. Available from: http://arxiv.org/abs/1512.03385.(Last access date 2021 Jun 01).
26. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211-252.
27. BeckAH Sangoi AR, Leung S, Marinelli RJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011;3: 108ra113.
28. Chang K, Collisson EA, Mills GB, et al. The cancer genome atlas-pan-cancer analysis project. Nat Genet 2013;45:1113-1120.
29. Saad ED, Spaffett P, Burzykowski T, et al. Disease-free survival as a surrogate for overall survival in cancer patients with HER2-positive, early breast cancer in trials of adjuvant trastuzumab for up to 1 year: a systematic review and meta-analysis. Lancet Oncol 2019;20: 361-370.
30. Dol V, Elberichi Z. Deep learning methods for mitosis detection in breast cancer histopathological images: A comprehensive review BT. In: Holtinger A, Goebel R, Mølgen M, Müller H, eds. Artificial Intelligence and Machine Learning for Digital Pathology: State-of-the-Art and Future Challenges. Cham: Springer International Publishing; 2020. p. 279-306.
31. Stenman S, Bychkov D, Kaucik H, et al. Antibody supervised training of a deep learning based algorithm for leukocyte segmentation in papillary thyroid carcinoma. IEEE J Biomed Health Inform 2021;25:422-428.
32. Turkki R, Linder N, Kovanen P, Pellinen T, Lundin J. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. J Pathol Inform 2016;7:38.
33. Linder N, Taylor JC, Colling R, et al. Deep learning for detecting tumour-infiltrating lymphocytes in testicular germ cell tumours. J Clin Pathol 2019;72:157-164.
34. Couture HD, Williams LA, Geradts J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer 2018;4:30.
35. Arunachalam HB, Mishra R, Daescu O, et al. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. PLoS One 2019;14, e0210706.
36. Turkki R, Linder N, Holopainen T, et al. Assessment of tumour viability in human lung cancer xenografts with texture-based image analysis. J Clin Pathol 2015;68:614-621.